# Generalized Brownian motion from a logical point of view

Jörg Kampen

January 9, 2012

## Abstract

We describe generalized Brownian motion related to parabolic equation systems from a logical point of view, i.e., as a generalization of Anderson's random walk. The connection to classical spaces is based on the Loeb measure. It seems that the construction of Roux in [11] is the only attempt in the literature to define generalized Brownian motion related to parabolic systems with coupled second order terms, where Lamé's equation of elastic mechanics is considered as an example. In this paper we provide an exact construction from a logical point of view in a more general situation. A Feynman-Kac formula for generalized Brownian motion is derived which is a useful tool in order to design probabilistic algorithms for Cauchy problems and initial-boundary value (of a class of) parabolic systems as well as for stationary boundary problems of (a class of) elliptic equation systems. The article includes a selfcontained introduction into all tools of nonstandard analysis needed, and which can be read with a minimum knowledge of logic in order to make the results available to a wider audience.

## 1 Introduction

Parabolic systems of equations are ubiquitous in sciences: they describe fields in physics and mechanics such as displacement fields of solids in elasticity, are used for modelling reaction-diffusion in chemical and biological processes, and even in mathematical finance certain systems with interacting Itô and point processes are related to parabolic systems and therefore beyond the realm of scalar equations (cf. [3]). On the other hand, although the powerful concept of Brownian motion entered many branches of physics, finance, and engineering such as heat transfer, dispersion (cf. Einstein,[4]), electrostatics and equity markets (Bachelier, [1]), information theory and noise (Shannon, [13], Brillouin, [2]), and quantum mechanics (Nelson). However, except for some work related to fluid mechanics (probabilistic approaches to the Navier-Stokes equations based on Malliavin calculus such as in Mattingly) and more abstract articles related to parabolic systems with coupling of lower order terms it seems that the work of Roux is the only attempt to construct a generalized Brownian motion related

to parabolic systems with coupled terms (such as Lamé's equation) of second order in the sense that certain expectations of functions of generalized processes solves Cauchy problems of a certain class of parabolic systems (to start with). It seems that the construction of Roux in [11] is the only published attempt in this direction. However, it is clear that this construction is not in an exact measure theoretical sense. This may be a parallel to the situation of the classical Brownian motion where Winer's construction of Wiener in 1923 was the first exact functional analytic description of Brownian motion after it has been introduced as a mathematical object on a more heuristic level by Bachelier and Einstein a long time before. However, it is not even clear on a heuristic level whether the construction of Roux really works (in this context we may say that a construction is 'correct on a heuristic level' if it leads to an algorithmic scheme (a Monte-Carlo method) where there is (at least) experimental evidence of convergence in a probabilistic sense. Well it seems by no means clear whether the Landau remainder terms in equations (29) and (30) in [11] become small as the time step size of the scheme becomes small. In this paper we describe a different approach to the problem stated. It is based on a generalized $\theta$ function corresponding to stochastic processes on the $n$-dimensional torus $\mathbb{T}^n = S^1 \times \cdots \times S^1$ (note that the $\theta$-function on the circle $S^1$ is the analogue of the fundamental solution on $\mathbb{R}$). It seems easier to construct generalized Brownian motions in the context of non-standard analysis. Since this theory is not that well-known this article provides a self-contained introduction to the subject including elementary stochastic analysis such as non-standard definitions of stochastic integrals, non-standard densities and Itô formulas. The theory is quite appealing in this respect since proofs are easier once the framework is established. Moreover, a very modest knowledge of logic is required to follow our introduction into the subject. Note that the non-standard theory used here is a consequence of the Zermelo-Fränkel theory together with the axiom of choice (so-called ZFC). We do not need additional assumptions. Other theories with explicit infinitesimals may be used, for example topoi used in the conext of synthetic differential geometry or Connes functional analytic theory (called non-commutative geometry), where compact operators play the role of infinitesimals. In any case we may interpret the part of nonstandard analysis as a part of ZFC-theory, although heavy use of the axiom of choice maybe against the taste of some advocates of classical descriptions. From a logical point of view this taste does not matter as does Connes remark that dart play may be better described by his theory (without proof). Similar as intuitionist mathematics nonstandard theory has been criticised to be not productive. In the latter case this criticism is stated in the sense that nonstandard analysis does not lead to results which do not have their classical counterparts. Adherents to this view are asked to give a classical counterpart to the construction of generalized Brownian motion and Feynman-Kac formula described in this paper. We close this first section with a description of the problem of generalized Brownian motion related to parabolic systems in a classical context. The parabolic equation systems considered in this paper the quadratic form

$$\frac{\partial \mathbf{u}}{\partial t} = \sum_{i,j=1}^{n} A^{ij} \frac{\partial^2 \mathbf{u}}{\partial x_i \partial x_j} + \sum_{j=1}^{n} \left( B^j \nabla \right) \mathbf{u} + C\mathbf{u}. \tag{1}$$

(We use the quadratic form for simplicity). Here,

$$\mathbf{u} = (u_1, \cdots, u_n)^T \qquad (2)$$

is a vector-valued function where for each $x \in \mathbb{R}^n$ and each $1 \leq i, j \leq n$

$$x \to A^{ij}(x) := \begin{pmatrix} a_{11}^{ij}(x) \cdots a_{1n}^{ij}(x) \\ \vdots \qquad \vdots \\ a_{n1}^{ij}(x) \cdots a_{nn}^{ij}(x) \end{pmatrix} \qquad (3)$$

may be assumed to be bounded $C^\infty$. Similarly $x \to B^j(x), j = 1, \cdots, n$ and $x \to C(x)$ may be assumed to be a $n \times n$-matrix-valued function respectively with bounded $C^\infty$-entries. If certain ellipticity conditions are satisfied then equations of the form (1) are called parabolic. Such conditions can be found in many classical textbooks such as ([6]). In analogy to the construction of classical Brownian motion our construction of generalized Brownian motion is in relation to equation with constant coefficients. The generalization to processes related to parabolic systems is quite straightforward. So let us assume for a moment that $a_{kl}^{ij}(x) \equiv a_{kl}^{ij}$. What we need for our construction is global existence and the requirement that for all $1 \leq i, j \leq n$ the matrices

$$A_\alpha = (A_\alpha^{ij}) := \left( \sum_{kl} a_{kl}^{ij} \alpha_k \alpha_l \right) \qquad (4)$$

are strictly elliptic for all $\alpha$ where all $\alpha_i \neq 0$.

Now in the case of scalar equations there is a natural correspondence between solutions of certain parabolic Cauchy problems and stochastic diffusion processes. We may put it this way: the fundamental solution of a scalar parabolic equation of type

$$\frac{\partial u}{\partial t} = \sum_{i,j=1}^n a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i \frac{\partial u}{\partial x_i} \qquad (5)$$

equals the transition density of a diffusion process of the form

$$dX_t = \sum_{i=1}^n b_i dt + \sum_{i,j=1}^n \sigma_{ij} dW_j, \qquad (6)$$

where $W = (W_1, \cdots, W_n)$ is a standard Brownian motion, and

$$(a_{ij}) = \sigma \sigma^T. \qquad (7)$$

A decomposition as in (7) exists if a certain regularity requirement is satisfied. More precisely the fundamental solution of (5) is the solution of the family of Cauchy problems

$$\begin{cases} \frac{\partial p}{\partial t} = \sum_{i,j=1}^n a_{ij} \frac{\partial^2 p}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i \frac{\partial p}{\partial x_i} \\ \\ p(0, x, y) = \delta_y(x), \end{cases} \qquad (8)$$

where for each $y \in \mathbb{R}^n$ $\delta_y(x) = \delta(y - x)$ along with the Dirac delta distribution $\delta$. On the other hand, if for each $x \in \mathbb{R}^n$ the stochastic differential equation

3

([6](#)) with initial data $X_0^x = x$ has a strong solution $X = (X_t^x)_{0 \leq t < \infty}$ which is associated with the Markov family $(X, \Omega, \mathcal{F} = (\mathcal{F})_{0 \leq t < \infty}, P^x)_{x \in \mathbb{R}^n}$, then the density $p$ should satisfy

$$P^x\left(X_t^x \in dy\right) = p(t, x, y)dy. \tag{9}$$

For example, the transition density related to the the family of Brownian motions $(W^x)_{x \in \mathbb{R}^n}$ is the fundamental solution of the heat equation. This relationship between stochastic processes and partial differential equations is quite useful as it provides us with probabilistic algorithm for the solution of Cauchy problems and boundary value problems. Moreover, analytic approximations of densities can be used in order to improve such probabilistic schemes (cf. ([7])). Our question is: Is there a analogous relationship for parabolic systems? Note that there are construction of fundamental solutions for parabolic systems such as (1) by the parametrix method. How does a stochastic process look like which satisfies a relation analogous to (9)? It seems that the paper of Roux (cf. [11]) is the only paper which has posed this question. However, as we have said, it seems that the definition given there needs some mathematical clarification since the Landau terms in equations (29) and equation (30) are not estimated. Maybe the situation is similar as with Bachelier's and Einstein's early work on Brownian motion. Here the work of Wiener in [14] provided the first exact description of the Wiener measure. The heat equation related to Brownian motion has constant second order coefficients. Hence, similar as in the scalar case we may consider parabolic systems with constant coefficients. Well the probabilistic interpretation of the fundamental solution is far less obvious. However, it is a start. As an example, let us consider Lamé's equation describing displacement fields in elasticity. In its time- dependent version it is

$$\frac{\partial \mathbf{v}}{\partial t} = [1 - \nu(n-1)] \nabla^2 \mathbf{v} + \nabla [\nabla \cdot \mathbf{v}], \tag{10}$$

and a typical problem is to solve it in some bounded domain $D = (0, T) \times \Omega$ with $\Omega \subset \mathbb{R}^n$, and where

$$\mathbf{v}(t, x) = B(x) \text{ for all } x \in \partial_p D$$

$$\mathbf{v}(0, x) = A(x) \text{ for all } x \in \Omega. \tag{11}$$

Here, $\partial_p D$ is the parabolic boundary of $D$ and the initial and boundary condition fields $A$ and $B$ will satisfy some compatibility conditions. In coordinates (10) reads (we use Einstein summation)

$$\frac{\partial v_i}{\partial t} = \sum_j [1 - \nu(n-1)] v_{i,jj} + v_{j,ji} \tag{12}$$

and in case of dimension $n = 2$ we see, with $a = [1 - \nu(n-1)]$ we have the representation

$$\begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} = \begin{pmatrix} a+1 & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} v_{1,11} \\ v_{2,11} \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} v_{1,12} \\ v_{2,12} \end{pmatrix}$$

$$+ \begin{pmatrix} a & 0 \\ 0 & a+1 \end{pmatrix} \begin{pmatrix} v_{1,22} \\ v_{2,22.} \end{pmatrix} \tag{13}$$

In this case our ellipticity assumption (4) amounts to the assumption that for all multiindices $\alpha \in \mathbb{Z}^n$ the matrix

$$\begin{pmatrix} \alpha_1^2(a+1) + a\alpha_2^2 & \alpha_1\alpha_2 \\ \alpha_1\alpha_2 & a\alpha_1^2 + (a+1)\alpha_2^2 \end{pmatrix} \tag{14}$$

is strictly elliptic (which is certainly true for $a > 0$). Note that this example is of the form (1) along with $B = 0$ and $C = 0$ (where $0$ denotes a matrix with zero entries in the former case and a vector with zero entries in the latter case. It is natural to define generalized Brownian motions to be processes which are related to the latter class of parabolic systems. Note that in elasticity stationary problems of the form

$$\begin{cases} \left[1 - \nu(n-1)\right] \nabla^2 \mathbf{v} + \nabla \left[\nabla \cdot \mathbf{v}\right] = 0 \text{ in } \Omega \\ \\ \mathbf{v} = \mathbf{g} \text{ for all } x \in \partial\Omega, \end{cases} \tag{15}$$

are of special interest. Here $\Omega \subset \mathbb{R}^n$ is some open domain. The probabilistic representation in terms of expectations with stopping times (first exit time) is quite appealing (also from a algorithmic point of view). In any case, the question is whether we find a a family of processes, say $B^{\otimes,A}$ (where $A$ encodes the information of constant coefficients $a_{kl}^{ij}$ of the diffusion), such that for $D = \mathbb{R}^n$ the function

$$(t,x) \rightarrow E^x \left( F \left( \mathbf{f}, B_t^{\otimes,A} \right) \right), \tag{16}$$

where $F$ is a rather simple functional with values in $\mathbb{R}^n$. We shall define $B_t^{\otimes,A}$ as an infinite vector of $n$ by $n$ matrices. Then $F$ will be just the infinite sum over all multiindices $\alpha$ of products of each $n$ by $n$ matrix entry of $B_t^{\otimes,A}$ with the vector entries of the infinite vector of vectors of the form

$$\left( \sum_{i=1}^{n} \hat{f}_{i\alpha} \exp\left(i2\pi\alpha x\right) \mathbf{e}_i \right)_{\alpha \in \mathbb{Z}^n} \tag{17}$$

encoding the information of the initial data $\mathbf{f}$. Here, $\hat{f}_{i\alpha}$ is the $\alpha$th Fourier coefficient vector of the $i$th component $f_i$ of the vector-valued function $\mathbf{f}$ and $\mathbf{e}_i$ denotes the $i$th unit basis vector of $\mathbb{R}^n$ (cf. next section for more details of this definition). Although it seems possible do to the construction in a classical framework of Wiener measures the nonstandard construction seems to be easier in this context. The connection to standard spaces is via the Loeb measure. Therefore, in the next section we shall introduce Anderson's random walk and the Loeb measure and make a precise definition of the generalized Brownian motion.

## 2 Anderson random walk and generalized Anderson random walk

The following construction may be put into a more classical framework, but the formulation seems more simple to me in the nonstandard framework. This is a matter of taste to some extent. In any case, from a logical point of view, we are working in ZFC. Moreover, we shall project to classical space finally. The

transition from nonstandard probability spaces to standard probability is via the Loeb measure. Let us recall the idea of the Loeb measure first (Readers with no background in nonstandard analysis are advised to read our selfcontained introduction into the subject starting with the next section first). Let $H$ be an hyperfinite, and let $P_I(H)$ be the set of all internal subset of $H$. Define

$$\mu : P_I(H) \to [0,1]^*$$
$$\mu(S) = \frac{|S|}{|H|}, \tag{18}$$

where $|.|$ denotes the cardinality of a hyperfinite set. The values of $\mu$ are in $[0,1]^*$ because subsets of a hyperfinite set $H$ have a smaller internal cardinality than $H$. Loeb observed that the map

$$\mu_L : P_I(H) \to [0,1]$$
$$\mu_L(S) = \mathrm{sh}\left(\frac{|S|}{|H|}\right), \tag{19}$$

is a measure, i.e., the map $\mu_L$ satisfies the countable additivity axiom. This is due to the fact that a family $(S_i)_{i \in \mathbb{N}}$ of mutually disjoint elements $S_i \in P_I(H)$ with

$$S := \cup_{i \in \mathbb{N}} S_i \in P_I(H) \tag{20}$$

is a finite actually, i.e.

$$S = S_1 \cup \cdots \cup S_k \tag{21}$$

for some $k \in \mathbb{N}$ (otherwise $S$ would be external). Next let us recall the main theorem concerning the Loeb measure. Let $(\Omega, \mathcal{A}, P)$ be an internal, finitely additive probability space, i.e.

i) $\Omega$ internal

ii) $\mathcal{A}$ is internal subalgebra of $\mathcal{P}(\Omega)$

iii) $P : \mathcal{A} \to {}^*\mathbb{R}$ is an internal function such that

iv) $P(\oslash) = 0$, $P(\Omega) = 1$, $\forall A, B : P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

The following theorem is the main theorem of non standard probability theory.

**Theorem 1.** *There is a standard ($\sigma$-additive) probability space $(\Omega, \mathcal{A}_L, P_L)$ such that*

*i) $\mathcal{A}_L$ is a $\sigma$-algebra with $\mathcal{A} \subseteq \mathcal{A}_L \subseteq \mathcal{P}(\Omega)$*

*ii) $P_L =^\circ P$ on $\mathcal{A}$*

*iii) For every $A \in \mathcal{A}_L$ and standard $\epsilon > 0$ there are $\mathcal{A}_i$ and $\mathcal{A}_o$ in $\mathcal{A}$ such that $A_i \subseteq A \subseteq A_o$ and $P(A_o \setminus A_i) < \epsilon$*

*iv) For every $A \in \mathcal{A}_L$ there is $B \in \mathcal{A}$ such that $P_L(A \Delta B) = 0$*

*The space $(\Omega, \mathcal{A}_L, P_L)$ is called a Loeb probability space*

In order to introduce stochastic processes we consider hyperfinite timelines

$$T = \{0, t_1, t_2, \cdots, t_N\}, \tag{22}$$

where $N$ is an infinite integer and $t_{i+1} - t_i$ are infinitesimal for all $i \in \{1, 2, \cdots, N\}$. In the following we assume that we have $\Delta t = t_{i+1} - t_i$ for all $i \in \{1, \cdots, N\}$, i.e., the discretization is uniform. An internal stochastic process is an internal map

$$X : T \times \Omega \to {}^*\mathbb{R}. \tag{23}$$

We may assume that it is adapted to a certain filtration of the Loeb algebra. We may then model Anderson's random walk (the nonstandard counterpart of Brownian motion) as an internal map

$$B : T \times \Omega \to {}^*\mathbb{R}$$

$$B(t, \omega) = \sum_{s<t} \omega(s)\sqrt{\Delta t}, \tag{24}$$

where we may model

$$\Omega = \{\omega : T \to \{-1, 1\}\} | \omega \text{ is internal}\} \tag{25}$$

It is clear how the associated probability measure and the Loeb filtration looks like in this case. It is well known that standard Brownian motion is just the shadow of this process, i.e., we may define

$$W(t, \omega) = {}^\circ B(t^-, \omega), \tag{26}$$

where $t^-$ is just the largest element in $T$ smaller or equal to $t$. Note that in (26) we defined the Brownian motion just in case of dimension $n = 1$. Generalization to higher dimension is straightforward, and we do not indicate the dimension when we use the symbol $W$ in the following. As usual processes are often written as families of random variables in the form $W_t$ or $X_t$ with the subscript $t$, and we adopt this convention. It will be clear from the context or irrelevant. This may be the most simple exact definition of a Brownian motion available. Note that Levy's theorem can be proved quite easily in this framework. We may construct more complicated processes from this easily, and we shall consider more general internal martingales later, but for the moment let us look at the Feynman-Kac formula in a very simple form. Consider the solution of the Cauchy problem on $[0, \infty) \times \mathbb{R}^n$

$$\begin{cases} \frac{\partial u}{\partial t} = \sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2} \\ u(0, .) = f, \end{cases} \tag{27}$$

where $f \in H^s(\mathbb{R}^n)$ for arbitrary $s \in \mathbb{R}$ (where $H^s$ is the usual Sobolev Hilbert space)- especially this means that $f$ is smooth and decays rather rapidly at infinity. We know that $u$ has the representation

$$u(t, x) = E^x (f(W_t)). \tag{28}$$

where $E^x$ is the expectation and the superscript $x$ indicates that we let start the Brownian motion at $x \in \mathbb{R}^n$. Similarly, if $v : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ solves the equation

$$\sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2} = f \text{ on } \Omega, \tag{29}$$

then $v$ has the representation

$$u(t,x) = E^x \left( f(W_{\tau_\Omega}) \right), \tag{30}$$

where $\tau_\Omega$ denotes the first exit time from the domain $\Omega$. Similarly, in order to define a Wiener measure on the $n$-torus $\mathbb{T}^n$ we may consider the fundamental solution $\theta$ of the problem with periodic boundary conditions

$$\begin{cases} \frac{\partial u}{\partial t} = \sum_{j=1}^n \frac{\partial^2 u}{\partial x_j^2}, \\ u(0,.) = f, \ u(t, x + \mathbf{e}_i) = u(t,x) \text{ for all } 1 \le i \le n, \end{cases} \tag{31}$$

where $\mathbf{e}_i$ denotes the $i$th vector of the standard basis of $\mathbb{R}^n$. This $\theta$-function is given for $t > 0$ by

$$\theta(t,x) = \sum_{\alpha \in \mathbb{Z}^n} \exp\left(2\pi i \alpha x - 4\pi \alpha_i^2 t\right) \tag{32}$$

where the sum is over all multiindices $\alpha = (\alpha_1, \cdots, \alpha_n)$ with entries $\alpha_i$ in the integers $\mathbb{Z}$. Note that for $\theta(t,.)$ converges in distributive sense to the $\delta$-distribution as $t \downarrow 0$. Accordingly,

$$u(t,x) = \int_{\mathbb{T}^n} f(y)\theta(t, x-y)dy. \tag{33}$$

Similar formulas hold for $n$-tori of any radius $R$ of course, and we may define associated Wiener measure $W^{\mathbb{T}_R^n}$ in the usual manner. As $R \uparrow \infty$ we get the standard Wiener measure. For elliptic problems such as (29) on bounded domains $\Omega$ we find equivalent representation of the form (34) or of the form

$$v(t,x) = E^x \left( f(W_{\tau_\Omega}^{\mathbb{T}_R^n}) \right), \tag{34}$$

for $R$ large enough. We shall consider measure which are defined on the $n$-torus. They are easily defined in the framework of nonstandard analysis and they lead to a description of generalized Brownian motion as we point out next. From the point of view of nonstandard analysis we may consider the functions $u, f$ to be standard parts of internal functions which we denote with the same symbols $u$ and $f$ for the sake of simplicity of notation. Let the time $t$ be the standard part of a time $t_M \in T$. If we observe the process $B^x$, i.e., the Anderson random walk starting at $x$, up to time $t_M$, then we know the value of $\omega(s)$ up to time $t$ and we nothing about $\omega$ for $s > t$. We may consider the equivalence classes

$$\omega \sim \tilde{\omega} \text{ iff } \forall s < t : \omega(s) = \tilde{\omega}(s), \tag{35}$$

and denote the corresponding equivalence classes by $[\omega]_t$. We define

$$\Omega_M = \{[\omega]_{t_M} | \omega \in \Omega\} \tag{36}$$

and

$$P_n : \Omega_M \to [0,1]$$

$$P_n([\omega]_{t_M}) = \tfrac{1}{2^M} \text{ for all } [\omega]_{t_M} \in \Omega_M. \tag{37}$$

8

Furthermore let us define the random variable

$$B_{t_M}^x : \Omega \to^* \mathbb{R}$$

$$B_{t_M}^x([\omega]_{t_M}) := x + \sum_{s<t} \omega(s)\sqrt{\Delta t},$$

(38)

where $\omega$ is some internal function with $\omega \in [\omega]_{t_M}$ (this is well defined since we get the same result for all $\omega \in [\omega]_{t_M}$ by definition of the equivalence relation $[]_{t_M}$). Next we consider a hyperfinite discretization of $^*\mathbb{R}$. Let $\Delta x$ be an infinitesimal hyperreal number and define

$$^*\mathbb{R}_{\Delta x} := \{k\Delta x | k \in^* \mathbb{Z}\},$$

(39)

where $^*\mathbb{Z}$ denotes the set of hyperintegers.

This discretization has the advantage that values of integrals in classical calculus are the standard parts of hyperfinite sums. Then we may introduce the density function $p$ defined on $T \times^* \mathbb{R}_{\Delta x} \times^* \mathbb{R}_{\Delta x}$ by

$$p(t_M, x, y) := \sum_{[\omega]_{t_M} \in \Omega_M} \delta_y \left(B_{t_M}^x([\omega]_{t_M})\right) P([\omega]_{t_M}),$$

(40)

where for each $y \in^* \mathbb{R}_{\Delta x}$

$$\delta_y(z) := \begin{cases} 1 \text{ iff } z = y \\ \\ 0 \text{ iff } z \neq y \end{cases}$$

(41)

denotes a hyperfinite Kronecker delta translated by $y$. Note the difference to classical calculus where the density is defined by a Cauchy problem with a delta distribution as initial data. The standard part of the function $p$ can be computed as a limit similar as in the nonstandard proof of the central limit theorem below. Furthermore, the internal function $u$ has a representation

$$u(t_M, x) = \sum_{y \in^* \mathbb{R}} \sum_{[\omega]_{t_M} \in \Omega_M} f(y)\delta_y \left(B_{t_M}^x([\omega]_{t_M})\right) P([\omega]_{t_M}),$$

(42)

and this representation may be used to get another proof of the Feynman-Kac formula. These representations of the density $p$ and the value function $u$ motivate analogous definitions in the context of the linear parabolic systems considered in the introduction. In order to do this we first consider an Anderson random walk in dimension $n$. We define

$$\Omega_n = \{\omega : T \to \{-1, 1\}^n | \omega \text{ internal }\},$$

(43)

and define

$$B : T \times \Omega_n \to^* \mathbb{R}^n$$

$$B(t, \omega) = \sum_{i=1}^n \left(\sum_{s<t} \omega_i(s)\sqrt{\Delta t}\right) \mathbf{e}_i$$

(44)

where $\omega_i$ is the $i$th component of the function $\omega$, and $\mathbf{e}_i$ is the $i$th element of the standard basis in $^*\mathbb{R}^n$. Next for any positive matrix $A$ we consider its

representation $A = Q\Lambda Q^T$ with $\Lambda = \text{diag}(\lambda_i)$ the diagonal matrix with diagonal entries $\lambda_i > 0$ and define

$$B^{\mathbf{\Lambda}} : T \times \Omega_n \rightarrow^* \mathbb{R}^n \otimes^* \mathbb{R}^n$$

$$B^{\mathbf{\Lambda}}(t,\omega)_{ij} = \sum_{i=1}^{n} \left( \sum_{s<t} \omega_i(s)\sqrt{\Delta t} \right) \lambda_i \delta_{ij}. \tag{45}$$

with the Kronecker $\delta$-function $\delta_{ij}$ (the subscript $ij$ indicates that the entry of the $i$th row and the $j$th clumn is defined). Furthermore, we define

$$B^{\mathbf{A}}(t,\omega) = QB^{\mathbf{\Lambda}}(t,\omega)Q^T \tag{46}$$

Generalized Brownian motions related to linear parabolic systems will be defined by an infinite vector of such matrix-valued Anderson random walks, where each entry encodes information of the diffusion coefficients $a_{kl}^{ij}$. The generalized Brownian motion we are going to construct can be represented by an infinite vector of $n$ by $n$ matrices. This may be described analogously to the so-called Kronecker description of tensor products but we do not go into this. For some total ordering of multiindices $\alpha \in \mathbb{Z}^n$ we may interpret $\mathbf{B}_{\mathbb{Z}^n} = (\mathbf{B}_\alpha)_{\alpha \in \mathbb{Z}^n}$ to be some infinite vector of $n$ by $n$ matrices matrices with entries in the complex numbers $\mathbb{C}$ and $\mathbf{V}_{\mathbb{Z}^n} = (\mathbf{v}_\alpha)_{\alpha \in \mathbb{Z}^n}$ some infinite vector of $n$-dimensional vectors with entries in the complex numbers. Then we define

$$\langle \mathbf{B}_{\mathbb{Z}^n}, \mathbf{V}_{\mathbb{Z}^n} \rangle := \sum_\alpha B_\alpha C_\alpha. \tag{47}$$

Moreover, the components of the product are denoted by

$$\langle \mathbf{B}_{\mathbb{Z}^n}, \mathbf{V}_{\mathbb{Z}^n} \rangle_i := \sum_\alpha (B_\alpha C_\alpha)_i, \tag{48}$$

and where $(B_\alpha C_\alpha)_i$ denotes the $i$th entry of the vector $B_\alpha C_\alpha$. Let $A_\alpha = (A_\alpha^{ij}) := \left( \sum_{kl} a_{kl}^{ij} 4\pi^2 \alpha_k \alpha_l \right)$. Then for $A_\alpha = Q_\alpha \Lambda_\alpha Q_\alpha^T$ for diagonal $\Lambda_\alpha$ with positive entries $\lambda_i > 0$ and orthogonal $Q$ define (with $\Delta x = \sqrt{\Delta t}$

$$B^{\sqrt{\Lambda}} : T \times \Omega_n \rightarrow^* \mathbb{R}_{\Delta x}^n \otimes^* \mathbb{R}_{\Delta x}^n$$

$$\left( B^{\sqrt{\Lambda}}(t,\omega) \right)_{ij} = \sum_{i=1}^{n} \left( \sum_{s<t} \omega_i(s)\sqrt{\Delta t} \right) \sqrt{\lambda_i}\delta_{ij}, \tag{49}$$

where $\delta_{ij}$ denotes the classical Kronecker delta, and for a positive definite matrix $A$ with decomposition $Q\Lambda Q^T$ and $\Lambda = \text{diag}(\lambda_i)$ define

$$B^{\sqrt{A}} : T \times \Omega_n \rightarrow^* \mathbb{R}_{\Delta x}^n \otimes^* \mathbb{R}_{\Delta x}^n$$

$$B^{\sqrt{A}}(t,\omega) := QB^{\Lambda^{1/2}}(t,\omega)Q^T. \tag{50}$$

A crucial observation is that

$$\mathbf{\Theta}^A(t, x-y) \approx E\left[ \sum_{\alpha \in \mathbb{Z}^n} \exp\left( i\alpha B^{\sqrt{A_\alpha}}(t,.) \right) \overrightarrow{\exp}\left( i2\pi\alpha(x-y) \right) \right] \tag{51}$$

where $\overrightarrow{\exp}\left(i2\pi\alpha(x-y)\right)$ denotes the $n$-dimensional vector with $n$ identical entries $\exp\left(i2\pi\alpha(x-y)\right)$. Note that (??) may be rewritten with

$$\overrightarrow{\delta}_{\infty}\left(x-y\right) := \left(\overrightarrow{\exp}\left(i2\pi\alpha(x-y)\right)\right)_{\alpha\in\mathbb{Z}^n}, \tag{52}$$

such that the resulting expression reminds of the fundamental solution. Note that the sums in (51) and (52) are standard sums over the standard set $\mathbb{Z}^n$. These are external objects. However, we are interested in certain projections on standard space related to classical solutions of parabolic systems. This motivates the definition

$$B^{\otimes,A}(t,.) := \left(\exp\left(i\alpha B^{\sqrt{A_\alpha}}(t,.)\right)\right)_{\alpha\in\mathbb{Z}^n} \tag{53}$$

For each $\alpha \in \mathbb{Z}^n$ we call $B^{\sqrt{A_\alpha}}$ the $\alpha$th mode of $B^{\otimes,A}$ Next we can state the main theorems of this paper.

**Theorem 2.** *(Feynman-Kac formula for parabolic systems on the n-torus) Let* $A = A^{ij}$, $A^{ij} = (a^{ij}_{kl})$ *with constant entries* $a^{ij}_{kl}$, *and assume that all* $A_\alpha$ *defined as above are strictly positive whenever* $\alpha = (\alpha_1, \cdots, \alpha_n)$ *where* $\alpha_i \neq 0$. *Let* $\mathbf{f} \in \left[L^2\left(\mathbb{T}^n\right)\right]^n$, *i.e.,* $\mathbf{f} = (f_1, \cdots, f_n)$ *along with* $f_i \in L^2\left(\mathbb{T}^n\right)$. *Put*

$$\mathbf{u}(t,x) =^\circ E\left[\sum_{\alpha\in\mathbb{Z}^n}\exp\left(i\alpha B^{\sqrt{A_\alpha}}(t,.)\right)\hat{\mathbf{f}}_\alpha\right], \tag{54}$$

*where*

$$\hat{\mathbf{f}}_\alpha = \left[\mathbf{f}_\alpha \overrightarrow{\exp}\left(i2\pi\alpha(x)\right)\right] := \tag{55}$$

$$\left(f_{1\alpha}\exp\left(i2\pi\alpha(x)\right),\cdots,f_{n\alpha}\exp\left(i2\pi\alpha(x)\right)\right)^T$$

*along with*

$$f_\alpha = (f_{1\alpha},\cdots,f_{n\alpha})^T, \tag{56}$$

$$f_{i\alpha} = \int_{\mathbb{T}^n}f_i(y)\exp\left(-i2\pi\alpha y\right)dy, \tag{57}$$

*and*

$$f_i(x) = \sum_{\alpha\in\mathbb{Z}^n}f_{i\alpha}\exp\left(i2\pi\alpha x\right). \tag{58}$$

*Or, written alternatively, put*

$$\mathbf{u}(t,x) =^\circ E\left[\left\langle B^{\otimes,A}(t,.), \overrightarrow{\mathbf{f}}_\infty\left(x\right)\right\rangle\right], \tag{59}$$

*where*

$$\overrightarrow{\mathbf{f}}_\infty\left(x\right) := \left(\mathbf{f}_\alpha \overrightarrow{\exp}\left(i2\pi\alpha(x)\right)\right)_{\alpha\in\mathbb{Z}^n}. \tag{60}$$

*Then* $\mathbf{u}$ *satisfies*

$$\begin{cases} \frac{\partial\mathbf{u}}{\partial t} = \sum_{i,j=1}^n A^{ij}\frac{\partial^2\mathbf{u}}{\partial x_i\partial x_j} \\ \\ \mathbf{u} = \mathbf{f} \end{cases} \tag{61}$$

*Moreover, if* $\mathbf{u}$ *is the solution of the equation (61), then this solution has the representation (63) or (54).*

11

This formula can be extended to more general cases with $B \neq 0$ and $C \neq 0$.

**Remark 3.** *Note that our the Feynman-Kac formula is interesting from a computational point of view because all $\alpha$-modes can be computed parallel and the contribution of an $\alpha$-mode decreases exponentially as the size $\alpha = \sum_{i=1}^{n} |\alpha_i|$ of the corresponding multiindex $\alpha$ increases (for $t > 0$). Note furthermore that schemes proposed in [7], [8], and [5], may be generalized to the present situation.*

As we see the generalized Brownian motion (63) corresponds to parabolic systems with constant coefficients. A more general class of stochastic processes can be constructed from this by infinite stochastic differential equations (similar as scalar partial differential equations are correlated to finite stochastic differential equations). For elliptic boundary problems we say that the general Hunt condition is satisfied for $B^{\otimes,A}$ if the standard Hunt condition is satisfied for every $\alpha$-mode $B^{\sqrt{A_\alpha}}(t,.)$, i.e., that every semiploar set $B^{\sqrt{A_\alpha}}(t,.)$ is polar for $B^{\sqrt{A_\alpha}}(t,.)$. We have

**Theorem 4.** *(probabilistic solution for elliptic boundary value problems of parabolic systems) Let $A$ and $\mathbf{f}$ satisfy the same assumptions as in the preceding theorem. Assume that $\Omega \subseteq \mathbb{T}^n$. Assume that for $\mathbf{f} = (f_1, \cdots, f_n)^T$ and all $1 \leq i \leq n$ all functions $f_i$ are bounded continuous functions on $\partial\Omega$, the boundary of $\Omega$. Assume that the general Hunt condition is satisfied, and that $\mathbf{v}$ is solution of the elliptic boundary value problem*

$$\sum_{i,j=1}^{n} A^{ij} \frac{\partial^2 \mathbf{v}}{\partial x_i \partial x_j} = 0$$

$$\lim_{x \to y} \mathbf{v}(x) = \mathbf{f}(y) \text{ for all regular } y \in \partial D \ \partial\Omega.$$

(62)

*Then*

$$\mathbf{v}(x) =^\circ E\left[\left\langle B_{\tau_\Omega^x}^{\otimes,A}, \overrightarrow{\mathbf{f}}_\infty (x) \right\rangle\right],$$

(63)

*where $\tau_\Omega^x$ denotes the first exit time from the domain $\Omega$ if the process starts at $x \in \mathbb{T}^n$.*

The proof of the first theorem is in the last section of this article, and the second follows from the first by standard arguments. In order to make the paper better readable to analysts we next provide a selfcontained introduction to nonstandard (stochastic) analysis.

# 3 Some remarks about mathematical systems with explicit infinitesimals and their relation to classical mathematics

The $\epsilon$-$\delta$-$\Pi_2$-formulas of Weierstrass almost eliminated explicit references to infinities and infinitesimals from classical mathematics. However they can be re-introduced in the framework of functional analysis, where in the theory of non-commutative geometry (invented by Alain Connes) infinitesimals are just compact operators for example. Other rival theories with explicit infinitesimals are topoi in synthetic differential geometry. These topoi are different from classical topoi in general since a term non datur does not hold. Logically, nonstandard analysis can be formulated in the framework of the Zermelo-Fränkel

system with the axiom of choice. Therefore, if no additional assumption (additional generosity concerning the enlargements of the universe) is made, then nonstandard analysis is logically equivalent to classical analysis based on ZFC. Modern system of mathematics differ with respect to the treatment of infinity. The rules for infinite objects can also affect the basic logical structure. This is true not only for intuitionist mathematics, but also for may other topoi which are used in different branches of mathematics. For example Euclid's assumption that 'for any two points in the plane, either they are equal, or they determine a unique line' may be interpreted to mean that the real number system which represents the line is actually a field. However if we intr0duce infinitesimals axiomatically by requiring that the set

$$\text{Inf}(0) := \left\{ x \neq 0 | x^2 = 0 \right\} \tag{64}$$

is not empty, then the axiom of synthetic differential geometry that the map

$$\mathbb{R} \times \mathbb{R} \to \mathbb{R}^{\text{Inf}(0)}$$

$$(r, s) \to [\delta \to r + \delta s] \tag{65}$$

is invertible becomes false. Typically in such axiomatic systems the law of excluded middle is also not valid. This means that different treatments of infinity (implicit as in Euclid or more explicit as in modern systems) lead to different interpretations of intuition and space. Intuition or our naive common sense concept of space cannot be a guideline to choose a system because these concepts are to vague in order to make any choice preferable.

# 4 An introduction to elementary nonstandard analysis

We extend the set real numbers $\mathbb{R}$ to a certain set of so-called hyperreal numbers ${}^*\mathbb{R}$. ${}^*\mathbb{R}$ will contain a copy of $\mathbb{R}$ and the infinitesimals and much more. Speaking roughly, ${}^*\mathbb{R}$ contains equivalence classes of sequences of real numbers, where equivalence of two sequences is induced by a nonprincipal ultrafilter on the integers $\mathbb{N}$. Real infinitesimal numbers are just equivalence classes of sequences which converge to zero. So formally the outline of this chapter is as follows: we construct a nonprincipal ultrafilter $\mathcal{F}$. Then we consider the set of sequences of real numbers $\mathbb{R}^{\mathbb{N}} = \{f : \mathbb{N} \to \mathbb{R}\}$. We define the set of hyperreals by

$$ {}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/\mathcal{F}, \tag{66}$$

where for $f, f' \in \mathbb{R}^{\mathbb{N}}$

$$f \sim f' \text{ iff } \{n | f(n) = f'(n)\} \in \mathcal{F}. \tag{67}$$

This is the so-called ultra-power construction of the hyperreals. To understand it in detail we first go into the construction of nonprincipal ultrafilters.

**Remark 5.** *Why do we not just take $\mathbb{R}^{\mathbb{N}}$ in order to define ${}^*\mathbb{R}$? The reason is that it seems difficult to define operations such that $\mathbb{R}^{\mathbb{N}}$ has the structure of a field. Pointwise operations do not work, because $(2, 0, 2, 0 \cdots) \cdot (0, 2, 0, 2, \cdots) = (0, 0, 0, 0 \cdots)$.*

# 5 Nonprincipal Ultrafilters

Let $S$ be an nonempty set and let $\mathcal{P}(S) = \{A | A \subseteq S\}$ be the power set of $S$. We say that a nonempty set $\mathcal{F} \subset \mathcal{P}(S)$ is a filter on $S$ if it is closed with respect to intersection and supersets, i.e.

   i) $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$,

   ii) $A \subset B,\ A \in \mathcal{F},\ B \subset S \Rightarrow B \in \mathcal{F}$

Note that $\oslash \in \mathcal{F}$ implies that for all $A \subseteq S$ $A \in \mathcal{F}$. A filter $\mathcal{F}$ is called proper iff $\oslash \notin \mathcal{F}$. A proper filter $\mathcal{F}$ on $S$ is called an ultrafilter if

   iii) for all $A \subset S$ either $A$ or $A^c$ is in $\mathcal{F}$,

where $A^c = S \setminus A$ denotes the complement of $A$ in $S$. An example of a proper filter which is not an ultrafilter is the filter of co-finite sets on an infinite set $S$, i.e. $\mathcal{F}^{co} := \{A \subset S | S \setminus A \text{ is finite}\}$. If $B \subset S$ is a set then

$$\mathcal{F}^B = \{A | S \supset A \supset B\}$$

is called the principal ultrafilter generated by $B$. We call a filter principal if it is generated by a singleton, i.e. a set with one element. In that case we write $\mathcal{F}^x$ instead of $\mathcal{F}^{\{x\}}$ for simplicity of notation. For an ultrafilter $\mathcal{F}$ this is equivalent to defining a filter as principal if it is generated by a finite set. For if $\mathcal{F}$ is generated by $B = \{x_1, \cdots, x_n\}$ assume that $\{x_i\} \notin \mathcal{F}$. Then $S \setminus \{x_i\} \in \mathcal{F}$. If this holds for all $x_i$, then

$$\oslash = \cap_{i=1}^n S \setminus \{x_i\} \cap B \in \mathcal{F}.$$

Hence, since $\mathcal{F}$ is proper there is one $x \in B$ such that $\{x\} \in \mathcal{F}$. Hence

$$\mathcal{F} := \{A | A \supset \{x\}\}.$$

To put it otherwise, if a principal ultrafilter is generated by a finite set, then it is generated by an element of that finite set. We call an ultrafilter nonprincipal if it is not principal. But do there exist nonprincipal ultrafilters? The answer is given in the next section.

# 6 The Axiom of Choice implies the Existence of nonprincipal ultrafilters

Let $S$ be a infinite set. First we say that a collection $H \subseteq P(S)$ has the finite intersection property (f.i.p.) if for any $n \in \mathbb{N}$

$$A_1, \cdots, A_n \in H \Rightarrow \cap_{i=1}^n A_i \neq \oslash. \tag{68}$$

Let

$$\mathcal{F}^H := \{A \subseteq S | A \supseteq B_1 \cap \cdots \cap B_n \text{ for some } n \text{ and } B_i \in H\} \tag{69}$$

be the filter generated by $H$.

**Lemma 6.** *If $H$ has f.i.p. then $\mathcal{F}^H \subseteq P(S)$ can be extended to an ultrafilter on $S$.*

14

Proof. Consider the p.o. $(\mathcal{P}, \subseteq)$ of filters, where $\mathcal{P} = \{\mathcal{F} | \mathcal{F} \supseteq \mathcal{F}^H\}$. If $L$ is a chain in $(\mathcal{P}, \subseteq)$, then $K := \cup L \in \mathcal{P}$ (since $\cup L$ is a filter). Recall that

$$\cup L = \{x | \exists y \in L : x \in y\}. \tag{70}$$

Hence, by Zorn's lemma $(\mathcal{P}, \subseteq)$ has a maximal element, which we name by $K$ again. Note that $K$ is a maximal proper filter on $S$ (since $K$ has the finite intersection property $K$ is proper). We want to show that $K$ is an ultrafilter. If $K$ is not an ultrafilter, then there is $A \subseteq S$ such that neither $A$ nor $A^c$ is in $K$. Assume that $K \cup \{A\}$ does not have f.i.p., i.e. assume that for some $B_1, \cdots, B_n \in K$ we have $B_1 \cap \cdots B_n \cap A = \oslash$. Then

$$B_1 \cap \cdots \cap B_n \cap A^c = B_1 \cap \cdots B_n. \tag{71}$$

Hence, $K \cup \{A^c\}$ has f.i.p.. Since $K$ is maximal $A^c \in K$. Hence $K$ is an ultrafilter.

**Remark 7.** *We present here the classical introduction to nonstandard analysis. Most but not all working mathematicians accept the axiom of choice (AC). Within the framework of the classical Zermelo-Fraenkel axiomatic system the AC leads to the famous paradoxes first discovered by Banach and Tarski. In that framework you can construct a subset $A$ of the two dimensional sphere $S^2$ such that for each natural number $n$ there are $n$ rotations $R_1, \cdots, R_n$ on $S^2$ such that*

$$S^2 = R_1(A) \cup R_2(A) \cup \cdots \cup R_n(A). \tag{72}$$

*Hence it is not possible to assign a measure to $A$ in a reasonable way. Especially, a rotation invariant measure would have to assign the value $1$ $\frac{1}{2}$, $\frac{1}{3}$, etc. to the same set $A$! Therefore it is interesting to design a constructive nonstandard analysis. Another reason to consider constructive versions of nonstandard analysis is the fact that we have to impose the continuum hypothesis (CH)*

$$\aleph_1 = 2^{\aleph_0} \tag{73}$$

*in order to ensure the uniqueness of $^*\mathbb{R}$. However, CH is even independent of ZFC as was shown by Gödel (who constructed a model of ZFC∪CH 1939) and Cohen (who constructed a model of ZFC∪¬CH 1939).*

**Theorem 8.** *If $S$ is an infinite set, then there exists a nonprincipal ultrafilter on $S$.*

Proof. Define the Frechet filter

$$\mathcal{F}^{co} = \{A \subseteq S | A^c \text{ is finite }\}. \tag{74}$$

$\mathcal{F}^{co}$ is a proper filter (not an ultrafilter, since there exist $S_1, S_2$ both infinite such that $S_1 \cup S_2 = S$). However, $\mathcal{F}^{Fr}$ has the f.i.p., so can be extended to a filter $\mathcal{F} \supset \mathcal{F}^{Fr}$ by the preceding lemma. We observe that $\mathcal{F}$ is nonprincipal, for if $\mathcal{F}^x$ is a principal filter generated by $\{x\}$, then $S \setminus \{x\} \notin \mathcal{F}^x$ but $S \setminus \{x\} \in \mathcal{F}$. Hence $\mathcal{F} \neq \mathcal{F}^x$ for all $x$. Hence $\mathcal{F}$ is not principal.

# 7 Extended real numbers, infinitesimals and unlimited real numbers

We take a nonprincipal ultrafilter on $\mathbb{N}$ and define

$$^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/\mathcal{F} \tag{75}$$

where for $x, y \in \mathbb{R}^{\mathbb{N}}$

$$x \sim y :\Leftrightarrow \{i | x_i = y_i\} \in \mathcal{F}. \tag{76}$$

We embed $\mathbb{R}$ in $^*\mathbb{R}$ identifying $x \in \mathbb{R}$ with $[(x, x, \cdots, x, \cdots)]$, i.e. $x \in \mathbb{R}$ is represented by the equivalence class of the real number sequence sequence which is constantly $x$. We define extensions of standard operations

$$
\begin{aligned}
+ \quad &: {}^*\mathbb{R} \times {}^*\mathbb{R} \to {}^*\mathbb{R} \quad [x] + [y] := [(x_i + y_i)], \\
\cdot \quad &: {}^*\mathbb{R} \times {}^*\mathbb{R} \to {}^*\mathbb{R} \quad [x] \cdot [y] := [(x_i \cdot y_i)],
\end{aligned}
\tag{77}
$$

and

$$[x]^{-1} := [(y_i)] \quad y_i = \begin{cases} \frac{1}{x_i} & \text{,if } x_i \neq 0 \\ 0, & \text{else} \end{cases}. \tag{78}$$

Furthermore, we define extensions of standard relations by

$$< \subseteq {}^*\mathbb{R} \times {}^*\mathbb{R}, \quad [x] < [y] \text{ iff } \{n | x_n \leq y_n\} \in \mathcal{F}. \tag{79}$$

Similarly for $<, >, =$. Furthermore, let $0^* = [(0, 0 \cdots, 0, \cdots)]$ and $1^* = [(1, 1 \cdots, 1, \cdots)]$. We identify $0^*, 1^*$ with $0, 1$ respectively. We have

**Proposition 9.** $(^*\mathbb{R}, +, \cdot, -1, <)$ *is an ordered field with zero* $0$ *and unit* $1$.

Proof. First we show that $(^*\mathbb{R}, <)$ is a total order, i.e.

$$\forall x, y \in {}^*\mathbb{R}: \ [x] < [y] \text{ or } [x] = [y] \text{ or } [x] > [y] \tag{80}$$

Note that

$$\mathbb{N} = \{i | x_i < y_i\} \cup \{i | x_i = y_i\} \cup \{i | x_i > y_i\}. \tag{81}$$

Since $\mathcal{F}$ is an ultrafilter exactly one of the three sets is in $\mathcal{F}$. For pedagogical reasons let us be a little more explicit at this point. Let $A_< = \{i | x_i < y_i\}$, $A_= = \{i | x_i = y_i\}$, and $A_> = \{i | x_i > y_i\}$. If $A_< \in \mathcal{F}$ then $A_= \cup A_> \notin \mathcal{F}$. Neither $A_=$ nor $A_>$ can then be in $\mathcal{F}$, because $\mathcal{F}$ is proper. If $A_< \notin \mathcal{F}$, then $A_= \cup A_> \in \mathcal{F}$. If in addition to $A_< \notin \mathcal{F}$ $A_= \in \mathcal{F}$, then $A_> \notin \mathcal{F}$ (since $\mathcal{F}$ is proper). On the other hand, if $A_< \notin \mathcal{F}$ and $A_= \notin \mathcal{F}$, then $A_> \in \mathcal{F}$ or $A_= \cup A_< \in \mathcal{F}$. However, the latter is impossible, since $A_< \notin \mathcal{F}$ implies $A_<^c \in \mathcal{F}$ and $A_= \notin \mathcal{F}$ implies $A_=^c \in \mathcal{F}$, so $A_<^c \cap A_=^c = (A_< \cup A_=)^c \in \mathcal{F}$. Hence $A_< \cup A_= \notin \mathcal{F}$. Hence, we have either $x < y$ (if $A_< \in \mathcal{F}$) or $x = y$ (if $A_= \in \mathcal{F}$) or $x > y$ (if $A_> \in \mathcal{F}$). Next the field laws have to be checked. These follow immediately from congruency of the operations $+$ and $\cdot$ and the inverse operation. By definition, this means that the operations are well defined. We show this for $+$, i.e. we show that

$$[(x_i + y_i)] = [(x_i)] + [(y_i)]. \tag{82}$$

Now, if $[(\tilde{x}_i)] = [(x_i)]$ and $[(\tilde{y}_i)] = [(y_i)]$ then

$$\{i|\tilde{x}_i = x_i\} \in \mathcal{F} \text{ and } \{i|\tilde{y}_i = y_i\} \in \mathcal{F}. \tag{83}$$

Hence,

$$\{i|\tilde{x}_i = x_i \ \& \ \tilde{y}_i = y_i\} \in \mathcal{F}, \tag{84}$$

by the filter property of $\mathcal{F}$. Hence,

$$\{i|\tilde{x}_i = x_i \text{ and } \tilde{y}_i = y_i\} \subseteq \{i|\tilde{x}_i + \tilde{y}_i = x_i + y_i\} \in \mathcal{F}. \tag{85}$$

Similar for the other operations $\square$

So what is an infinitesimal? We can say an infinitesimal is a number in ${}^*\mathbb{R}$ which has modulus smaller than all standard real numbers. So let us define modulus first: for all $x = [(x_n)] \in {}^*\mathbb{R}$ we set

$$|x| = |[(x_n)]| := [(|x_n|)]. \tag{86}$$

and we say that $x \in {}^*\mathbb{R}$ is an infinitesimal if

$$\Phi_{\inf}(x) \equiv \forall \epsilon \in \mathbb{R}(0 < \epsilon \rightarrow |x| < \epsilon) \tag{87}$$

holds. Recall that we identified $\mathbb{R}$ with a subset of ${}^*\mathbb{R}$. A few remarks are in order. First, each standard real number $x \in \mathbb{R}$ has a whole bunch of infinitesimals around it. They are called the monad of $x$ and we have proved that they are linearly ordered. We define

$$\mathrm{md}(x) := \{y|\Phi_{\inf}(y - x)\}. \tag{88}$$

Well, we have not defined what $y - x$ is, but it should be clear from the context that

$$\forall x, y \in {}^*\mathbb{R} : y - x = [(y_i)] - [(x_i)] = [(y_i - x_i)]. \tag{89}$$

Since ${}^*\mathbb{R}$ is linearly ordered we may call $x \in {}^*\mathbb{R}$ positive infinitesimal iff $x > 0$ and $\Phi(x)$. An element $x \in {}^*\mathbb{R}$ is called unlimited if for all $r \in \mathbb{R} \ r < |x|$.

# 8 Further remarks

The ultrafilter construction of the hyperreal numbers is unique if the continuum hypothesis holds. Gödels proof of equiconsistency of $ZF$ and $ZF \cup AC \cup CH$ shows that this is not as bad a situation as one may think. Let us consider the question of cardinality of $R^*$ with and without this hypothesis. Since $\mathbb{R}$ can be embedded into ${}^*\mathbb{R}$ and ${}^*\mathbb{R}$ consists of equivalence classes of elements of $\mathbb{R}^{\mathbb{N}}$, it is clear that

$$\aleph_1 \leq \mathrm{card}\mathbb{R} = 2^{\aleph_0} \leq \mathrm{card}({}^*\mathbb{R}) \leq \mathrm{card}(\mathbb{R}^{\mathbb{N}}) = 2^{2^{\aleph_0}} \geq \aleph_2 \tag{90}$$

holds. Next we show that for each standard number $x \in \mathbb{R}$ the cardinality of the monad of $x$ is at least the cardinality of the real number. We construct an injection from the standard real numbers into the monad of 0 (the injections to $md(x)$ can be constructed in the same way). Consider representations of real numbers $r \in \mathbb{R}$ by decimal expansions

$$r = r_m r_{m-1} \cdots r_0.r_{-1} r_{-2} \cdots . \tag{91}$$

17

Consider an enumeration of the prime numbers $p_0, p_1, p_2 \cdots$ and the map

$$r \to [(r_n^+)] \tag{92}$$

with

$$r_n^+ := p_0^{-r_m} p_1^{-r_{m-1}} \cdots p_n^{-r_{m-n}}. \tag{93}$$

This is indeed an injection by uniqueness of prime factorization of natural numbers. If the (general) continuum hypothesis holds then the cardinality of $^*\mathbb{R}$ is $\aleph_2$ and $^*\mathbb{R}$ is unique.

The set $^*\mathbb{R}$ which we constructed as equivalence classes of real sequences with respect to an nonprincipal ultrafilter $\mathcal{F}$ in the last Chapter is just a set of numbers. To do nonstandard analysis and nonstandard stochastic analysis later, we need to talk about sets, functions and relations. It is possible to extend the ultrafilter construction to this kind of objects. First, for $A \subset \mathbb{R}$ we may define

$$x \in A^* \text{ iff } \{n | x_n \in \mathcal{F}\}. \tag{94}$$

If $\tilde{x} = x \in A^*$, then $\{n | \tilde{x}_n = x_n\} \in \mathcal{F}$. Hence,

$$\mathcal{F} \ni \{n | \tilde{x}_n = x_n \text{ and } x_n \in A\} \subseteq \{n | \tilde{x}_n \in A\} \in \mathcal{F}. \tag{95}$$

Similar, a function $f : \mathbb{R} \to \mathbb{R}$ extends to a function

$$f^* : {}^*\mathbb{R} \to {}^*\mathbb{R}, \quad f^*(x) = [(f(x_n))]. \tag{96}$$

This is well defined, again. If $x := [(x_n)] = [(y_n)] =: y$, then $\{n | x_n = y_n\} \in \mathcal{F}$. Hence, $\{n | x_n = y_n\} \subseteq \{n | f(x_n) = f(y_n)\} \in \mathcal{F}$. Hence $f(x) = f(y)$. Similarly, a relation $S \subseteq (\mathbb{R})^m$ extends to

$$S^* \subseteq ({}^*\mathbb{R})^m \quad (x_1, \cdots, x_m) \in S^* \text{ iff } \{n | (x_{1n}, \cdots, x_{mn}) \in S\} \in \mathcal{F} \tag{97}$$

So we could reconstruct all properties of analysis on $\mathbb{R}$ which are expressable in first order logic by the ultrafilter construction. However, the ultrafilter construction is at the very heart of Loos theorem which is essentially the transfer principle of nonstandard analysis. In order to state and proof the theorem of Loos we need some basics of mathematical logic, which we describe in the next Section.

# 9 Elements of Mathematical Logic (preparation for the theorem of Loos)

In order to fix notation for a proof of the thorem of Loos, we introduce basics of first order logic, i.e. the setup of formal first order languages, the logical axioms and rules of inference, and first order semantics.

## 9.1 Formal first order Languages

The object first order language which we consider consists of an alphabet with

   ai) logical symbols: $\neg$ (not) $\wedge$ (and) $\forall$ (for all) $=$ (equality)

aii) variables: elements of $\mathcal{V} := \{v_i | i \in \mathbb{N}\}$

aiii) relation symbols: elements of $\mathcal{R} := \{R_i | i \in I\}$

aiv) function symbols: elements of $\mathcal{F}_f := \{f_i | i \in J\}$

av) constants: elements of $\mathcal{C} := \{c_k | k \in K\}$

avi) auxiliary symbols: ) , (

Here, $I, J, K$ are arbitrary index sets. Next we build terms inductively as follows:

ti) all $v_i \in \mathcal{V}$ and $c_k \in \mathcal{C}$ are terms

tii) if $t_1, \cdots, t_{\mu(j)}$ are terms, and $j \in J$, then $f_j\left(t_1, \cdots, t_{\mu(j)}\right)$ is a term.

tiii) No other concatenations of symbols are terms

Here, $\mu : J \to \mathbb{N} \setminus \{0\}$ is a function which assigns to each index $j \in J$ the number of arguments of the function $f_j$. We denote the set of terms by Tm.
Next we build up formulas.

fi) if $t_1, t_2$ are formulas, then $t_1 = t_2$ is a formula

fii) if $t_1, \cdots, t_{\lambda(i)}$ are terms, then $R_i\left(t_1, \cdots, t_{\lambda(i)}\right)$ is a formula

fiii) if $\phi$ and $\psi$ are formulas, then so are $\neg\phi$, $(\phi \wedge \psi)$, and $\forall v_i \phi$ for some $v_i \in \mathcal{V}$.

fiv) No other strings of symbols are formulas

Here $\lambda : I \to \mathbb{N} \setminus \{0\}$ is a function which assign to each $i \in I$ the number of arguments of the relation $R_i$. For convenience and (partially) abusing these formal restrictions we fell free to use

coni) the symbols $u, v, w, x, y, z$ with and without indices as variables

conii) $\phi, \psi, \alpha, \beta\gamma$ with and without indices as symbols for formulas.

Furthermore we use the abbreviations

abi) $(\phi \vee \psi)$ stands for $\neg(\neg\phi \wedge \neg\psi)$

abii) $(\phi \to \psi)$ stands for $\neg(\phi \wedge \neg\psi)$

abiii) $(\phi \to \psi)$ stands for $(\phi \to \psi) \wedge (\psi \to \phi)$

abiv) $\exists\phi$ stands for $\neg\forall\neg\phi$

We also use the following conventions which make the formulas more accessible to the human reader

ci) $\wedge$ and $\vee$ have priority to $\to$ and $\leftrightarrow$

cii) $\neg$ has priority to $\vee$ and $\wedge$

ciii) $t_1 \neq t_2$ stands for $\neg t_1 = t_2$

civ) $\forall u, v, w \cdots$ stands for $\forall u \forall v \forall w$ and $\exists x, y, z \cdots$ stands for $\exists x \exists y \exists z$.

cv) we avoid the brackets if this does not lead to ambiguities, e.g. $\phi_1 \wedge \phi_2 \wedge \phi_3$ stands for $((\phi_1 \wedge \phi_2) \wedge \phi_3)$ etc.

Next we need the notion of free variables. We define the set of the free variables of a formula $\phi$ Free($\phi$) inductively. We set

Fi) Free($\phi$) = $\{v \in \mathcal{V} | v$ occurs in $\phi\}$

Fii) Free($\neg\phi$) = Free($\phi$)

Fiii) Free($\phi \wedge \psi$) = Free($\phi$) $\cup$ Free($\psi$)

Fiv) Free($\forall v \phi$) = Free($\forall v \phi$) $\setminus \{v\}$.

Furthermore, we need the syntactic operation of substitution of a variable $v$ in a string of symbols $\phi$ by a term $t$. Let $\phi\big|_t^v$ be the string of symbols which is obtained, if all occurrences of $v$ are replaced by $t$. If the variable $v$ in a formula $\psi$ is in the domain of a quantifier $\forall u$ and $u$ occurs in the term $t$, then the variable $u$ becomes part of the domain of quantor $\forall u$. If this does not happen for all variables in the term $t$, then $t$ is called free for $v$ in $\psi$. The formal language is determined by the functions $\lambda$, $\mu$, and the set $\mathcal{C}$. We write

$$L = (\lambda, \mu, \mathcal{C}) . \tag{98}$$

## 9.2 Logical axioms and rules of inference

There are a lot if axiomatic systems which lead to logical first order completeness. We use a very simple form. The logical axioms are

tl) tautologies of propositional calculus

ql) quantifier rules

$$\forall x \phi \to \phi\big|_t^x \text{ if } t \text{ is free for } x \text{ in } \phi$$
$$\forall x (\phi \to \psi) \to (\phi \to \forall x \psi) \text{ if } x \notin \text{Free}(\phi). \tag{99}$$

il) axioms of identity

$$x = x$$

$$x = y \to (x = z \to y = z)$$

$$x = y \to f_j(\cdots, x, \cdots) = f_j(\cdots, y, \cdots) \tag{100}$$
( same place of $x$ and $y$ in argument $\mu_j$- tuple of $f$)

$$x = y \to R_i(\cdots, x, \cdots) = R_i(\cdots, y, \cdots)$$
( same place of $x$ and $y$ in argument $\lambda_i$-tuple of $R_i$)

Next we have the rules of inference

$$\phi \to \psi, \ \phi \vdash \psi. \tag{101}$$

Finally, the rule of generalization is

$$\phi \vdash \forall x \phi. \tag{102}$$

## 9.3 First order semantics

Let $L = (\lambda, \mu, \mathcal{C})$ be a formal language as stated above. A $L$-structure

$$\mathcal{A} = \left( A, \left( R_i^A \right)_{i \in I}, (f_j^A)_{j \in J}, (c_k^A)_{k \in \mathcal{C}} \right) \tag{103}$$

is determined by the data of a set $A$ of individuals, a family of relations $R_i^A \subseteq A_{i \in I}^{\lambda(i)}$, a family of functions $f_j^A : A^{\mu(j)} \to A$, and a set of constants $\left\{ c_k^A \in A | k \in K \right\}$. A map

$$e : \mathcal{V} \to A \tag{104}$$

is called a valuation map. Furthermore, we define

$$e \big|_x^a (v) := \begin{cases} e(v) & \text{,if } v \neq x \\ a & \text{, if } v = x \end{cases}. \tag{105}$$

The value of a term $t \in \mathrm{Tm}$ with valuation map $e$ is defined inductively. We have

tei) $v^A(e) := e(v)$

teii) $c_k^A(e) := c_k^A$

teiii) $f_j \left( t_1, \cdots, t_{\mu(j)} \right)^A (e) = f_j^A \left( t_1^A(e), \cdots, t_{\mu(j)}^A(e) \right)$

The validity of a formula is a relation between the model $\mathcal{A}$, the formula $\phi$ and the valuation map $e$ which we write

$$\mathcal{A} \models \phi(e), \tag{106}$$

and define inductively

mati) $\mathcal{A} \models t_1 = t_2(e)$ iff $t_1^A(e) = t_2^A(e)$

matii) $\mathcal{A} \models R_i \left( t_1, \cdots, t_{\lambda(i)} \right)(e)$ iff $R_i^A \left( t_1^A(e), \cdots t_{\lambda(i)}^A(e) \right)$.

For connectives and quantifiers the following rules hold

mfi) $\mathcal{A} \models \neg \phi(e)$ iff $\mathcal{A} \not\models \phi(e)$

mfii) $\mathcal{A} \models (\phi \wedge \psi)(e)$ iff $\mathcal{A} \models \phi(e)$ and $\mathcal{A} \models \psi(e)$

mfiii) $\mathcal{A} \models \forall x \phi(e)$ iff $\mathcal{A} \models \phi(e\big|_a^x)$ for all $a \in A$.

### 9.4 The transfer principle

Transfer principle

$$^*\mathbb{R} \models \phi^*(e) \text{ iff } \mathbb{R} \models \phi(e). \tag{107}$$

$\phi^*$ is defined inductively. For constants $c \in C$ and variables $x \in \text{Var}$ we have

$$c^* = c \text{ and } x^* = x. \tag{108}$$

If $f$ is a k-ary function and $\tau_1, \cdots, \tau_k$ are terms, then

$$f(\tau_1, \cdots, \tau_k)^* = f^*(\tau_1^*, \cdots, \tau_k^*). \tag{109}$$

Similarly, if $P$ is a k-ary relation and $\tau_1, \cdots, \tau_k$ are terms, then

$$P(\tau_1, \cdots, \tau_k)^* = P^*(\tau_1^*, \cdots, \tau_k^*). \tag{110}$$

This defines $\phi^*$ for all atomic formulas $\phi$. Furthermore, if $\psi = \neg\phi$, then

$$\psi^* \equiv \neg\phi^*, \tag{111}$$

and if $\psi = (\phi_1 \to \phi_2)$, then

$$\psi^* = (\phi_1^* \to \phi_2^*). \tag{112}$$

It is well known that $\{\neg, \to\}$ is a complete set of connectives for classical propositional calculus. So, the $*$-transform for $\wedge$ etc. follow from the $*$-transform for $\neg$ and $\to$. Next, including quantifiers, if $\psi = \forall x \in A\phi$, then

$$\psi^* \equiv \forall x \in A^*\phi^*, \tag{113}$$

and the transform of $\psi = \exists x \in A\phi$ is implied by $\exists \equiv \neg\forall\neg$. The transfer principle is a consequence of the theorem of Loos.

## 10 The theorem of Loos

Let $S$ be an index set. For all $s \in S$ we consider the model

$$\mathcal{A}^s = (A^s, \mathrm{R}^s, \mathrm{F}^s, \mathrm{C}^s), \tag{114}$$

where $\mathrm{R}^s$ is a set of relations, $\mathrm{F}^s$ is a set of functions, $\mathrm{C}^s$ is a set of constants. We saw that the axiom of choice implies for any system $D$ of subsets of a set $M$ which is closed under finite intersections the existence of a superset $\mathcal{F} \supset D$ which is a nonprincipal ultrafilter. Next we define the ultraproduct of the $\mathcal{A}^s$ with respect to $\mathcal{F}$. Let us consider the set of sequences

$$\Pi_{s \in S} A^s = \{(a^s) \,|\, a^s \in A^s \text{ for all } s \in S\}. \tag{115}$$

Two sequences $(a_1^s)_{s \in S}$ and $(a_2^s)_{s \in S}$ are called equivalent with respect to $\mathcal{F}$ if they coincide on a set of $\mathcal{F}$, ie.

$$(a_1^s)_{s \in S} \sim_{\mathcal{F}} (a_2^s)_{s \in S} \text{ iff } \{s | a_1^s = a_2^s\} \in \mathcal{F}. \tag{116}$$

Since $\mathcal{F}$ is a filter the relation $\sim_{\mathcal{F}}$ is an equivalence relation. For the sake of brevity we shall write $(a_1^s)$ instead $(a_1^s)_{s \in S}$ in the following. Now let

$$A = \{[(a^s)] \,|\, (a^s) \in \Pi_{s \in S} A^s\}, \tag{117}$$

where

$$[(a^s)] = \{(b^s) | (a^s) \sim_{\mathcal{F}} (b^s)\}. \tag{118}$$

Next on $A$ we define the $R_i$ for $i \in I$, $F_j$ for $j \in J$, $C_k$ for $k \in K$ such that we can construct the structure

$$\mathcal{A} = \left(A, (R_i)_{i \in I}, (F_j)_{j \in J}, (c_k)_{k \in K}\right). \tag{119}$$

We call $\mathcal{A}$ the ultraproduct of the structures $\mathcal{A}^s$ and write

$$\mathcal{A} = \Pi_{s \in S} \mathcal{A}^s / \mathcal{F}. \tag{120}$$

The relations $R_i$ are defined as follows. We set

$$R_i\left([(a_1^s)], \cdots, \left[(a_{\lambda(i)}^s)\right]\right) \text{ iff } \left\{s \in S \,|\, R_i\left(a_1^s, \cdots, a_{\lambda(i)}^s\right)\right\} \in \mathcal{F}. \tag{121}$$

We define functions $F_j$ for $j \in J$ as follows:

$$f_j\left([(a_1)^s], \cdots, [(a_{\mu(j)})^s]\right) := \left[\left(f_j\left(a_1^s, \cdots, a_{\mu(j)}^s\right)\right)\right] \tag{122}$$

Similar for constants $c_k$

$$c_k = [(c_k^s)]. \tag{123}$$

In order to formulate the theorem of Loos we need the concept of valuation of sequences. For all $s \in S$ let $e^s$ be a valuation maps of the variables in $\mathcal{A}^s$. We define

$$[(e^s)] : \text{Vbl} \to A,$$
$$\tag{124}$$
$$[(e^s)](v_\nu) = [e^s(v_\nu)]$$

If $a = [(a^s)] \in A$ then

$$[(e^s)]\big|_a^x = \left[e^s\big|_a^x\right] \tag{125}$$

We abbreviate $e = [(e^s)]$. The theorem of Loos then says

**Theorem 10.**

$$A \models \phi(e) \text{ iff } \{s | A^s \models \phi(e^s)\} \in \mathcal{F} \tag{126}$$

Proof. The proof is by induction on the construction of the formulas $\phi$. If $\phi$ is a prime formula, then the equivalence follows by the equivalence of constants $c_k$ functions $f_j$ and relations $R_k$ as indicated before. Note also that

$$t^{\mathcal{A}}[(e^s)] = \left[\left(t^{\mathcal{A}^s} e^s\right)\right] \tag{127}$$

If $\phi \equiv \neg\psi$, then

$$\mathcal{A} \models \neg\psi(e) \quad \text{iff } \mathcal{A} \not\models \psi(e)$$

$$\text{iff } \mathcal{A} \not\models \psi\left([(e^s)]\right)$$

$$\text{iff } \{s | \mathcal{A}^s \models \psi(e^s)\} \notin \mathcal{F} \tag{128}$$

$$\text{iff } S \setminus \{s | \mathcal{A}^s \not\models \psi(e^s)\} \in \mathcal{F}$$

$$\text{iff } \{s | \mathcal{A}^s \not\models \psi(e^s)\} \in \mathcal{F}$$

If $\phi \equiv \psi \wedge \chi$, then

$$
\begin{aligned}
\mathcal{A} \models (\psi \wedge \chi)(e) \quad &\text{iff } \mathcal{A} \models \psi(e) \text{ and } \mathcal{A} \models \chi(e) \\
&\text{iff } \{s | \mathcal{A}^s \models \psi(e^s)\} \in \mathcal{F} \text{ and } \{s | \mathcal{A}^s \models \chi(e^s)\} \in \mathcal{F} \\
&\text{iff } \{s | \mathcal{A}^s \models \psi(e^s)\} \cap \{s | \mathcal{A}^s \models \chi(e^s)\} \in \mathcal{F} \\
&\text{iff } \{s | \mathcal{A}^s \models (\psi \wedge \chi)(e^s)\} \in \mathcal{F}
\end{aligned}
\tag{129}
$$

Finally, if $\phi$ is of the form $\forall x \psi$, then

$$
\begin{aligned}
\mathcal{A} \models \phi(e) \quad &\text{iff } \mathcal{A} \models \psi\left(\left[e^s\big|_a^x\right]\right) \text{ for all } a \in A \\
&\text{iff } \mathcal{A} \models \psi\left(\left[e^s\big|_{a^s}^x\right]\right) \text{ for all } (a^s) \in \Pi_{s \in S} A^s \\
&\text{iff } \left\{s | \mathcal{A}^s \models \psi\left(e^s\big|_{a^s}^x\right)\right\} \in \mathcal{F} \text{ for all } (a^s) \in \Pi_{s \in S} A^s \\
&\text{iff } \left\{s | \mathcal{A}^s \models \psi\left(e^s\big|_{a^s}^x\right) \text{ for all } a^s \in A^s\right\} \in \mathcal{F} \\
&\text{iff } \{s | \mathcal{A}^s \models \forall x \psi(e^s)\} \in \mathcal{F}
\end{aligned}
\tag{130}
$$

Here, the fourth step can be justified as follows. Let

$$
U = \left\{s | \mathcal{A}^s \models \psi\left(e^s\big|_{a^s}^x\right) \text{ for all } a^s \in A^s\right\}
\tag{131}
$$

For any sequence $(b^s) \in \Pi_{s \in S} A^s$ we have then On the other hand, if $U \notin \mathcal{F}$, then $s \setminus U \in \mathcal{F}$.

$$
V = \left\{s | \mathcal{A}^s \not\models \psi\left(e^s\big|_{a^s}^x\right) \text{ for some } a^s \in A^s\right\} \in \mathcal{F}
\tag{132}
$$

Then define a sequence $(b^s)$ where $b^s$ equals some $a^s$ with $\mathcal{A}^s \not\models \psi\left(e^s\big|_{a^s}^x\right)$ if $s \in V$ and some $a^s \in A^s$ else. Then

$$
V \subseteq \left\{s | \mathcal{A}^s \not\models \psi\left(e^s\big|_{b^s}^x\right)\right\} \notin \mathcal{F},
\tag{133}
$$

and since $V \in \mathcal{F}$ we have

$$
\left\{s | \mathcal{A}^s \models \psi\left(e^s\big|_{b^s}^x\right)\right\} \notin \mathcal{F},
\tag{134}
$$

which is impossible.

# 11 Introduction to enlarged universes and measure theory

Stochastic analysis is a part of measure theory and measures assign numbers to sets. Nonstandard stochastic analysis is applied Loeb measure theory and Loeb measures assign values to internal sets. So we have to agree on what is a set (which axioms of set theory we take for granted). Next a good framework of measure theory are universes and we have to define enlargements of universes,

which we do by the ultrafilter construction. Then we define internal sets within the enlarged universe. This is the program. Some remarks are in order.

The foundations of mathematics can be displayed either in the classical framework of set theory (especially the Zermelo-Fraenkel system) or within the framework of category theory (and here especially in the framework of topos theory). The categorical approach and especially the topos theoretic approach is the most advanced view on the foundations since it allows us to represent different point of views on mathematics under one uniform roof. E.g. the logic of topos theory is intrinsically intuitionist but sheafications lead to classical mathematics in an elegant way. However, since topos theory is not that popular we consider set theory in this introduction. Since the subject of nonstandard stochastic analysis is a challenge in itself we do not want to make things more difficult by using the categorical approach. However we do not use exactly ZFC. The reason is that restricted quantifiers are enough for our purposes. Recall the ontological premise of Quine ("On what there is"): "to be is to be value of a variable". Variables are exactly what we quantify over. So what exists in restricted ZFC (sometimes also called "bounded Zermelo") are elements of sets.

## 12  Bounded Zermelo

Bounded Zermelo is ZFC with quantification restricted to existing sets. The language of bounded Zermelo is a normal set theoretic language, with the exception of the restricted quantification rule, i.e. for any formula $\phi(x)$ with the free variable $x$ and sets $a, b$

$$\exists x \in a\, \phi(x), \ \forall x \in b\, \phi(x) \tag{135}$$

are formulas. Denote the usual set theoretic language with bounded quantifiers by $\mathcal{L}_R$. We list the axioms of restricted ZFC (RZFC in symbols).

(E) (Extensionality) $y = x$ iff, for all $z$, $z \in x$ iff $z \in y$.

(RC) (Restricted Comprehension) If $\phi(x)$ is an $\mathcal{L}_R$-formula (with quantifiers restricted) and free variable $x$ and $a$ is a set, then there exists a set $b$ with $x \in b$ iff $x \in a$ and $\phi(x)$.

(NS) (Null Set) There exists a set $\oslash$ such that for all $x$ $x \notin \oslash$.

(P) (Pair) For all $x$ and $y$ there exists $z$ with $u \in z$ iff $u = x$ or $u = y$.

(U) (Union) For all $x$ there exists $y$ with $z \in y$ iff there exists $w$ with $z \in w \in y$.

(PS) (Power Set) for all $x$ there exists $y$ with $z \in y$ iff $z \subseteq x$. The set $y$ will be denoted by $P(x)$

(F) (Foundation) For all $x \neq \oslash$ there exists a set $y \in x$ with $y \cap x = \oslash$.

(I) (Axiom of Infinity) There exists a set $\mathbb{N}$ such that $\oslash \in \mathbb{N}$ and $x \in \mathbb{N}$ implies $x \cup \{x\} \in \mathbb{N}$

(AC) (Axiom of Choice) if $I \neq \oslash$ is an index set and for all $i \in I$ $X_i \neq \oslash$, then $\Pi_{i \in I} X_i \neq \oslash$.

Bounded Zermelo is well known to be equiconsistent to a model of the first order theory of well-pointed topoi. However, as explained before, we are not dealing with the topos-theoretic view here.

# 13   Mathematics within first order logic

The goal of the present chapter is to extend the construction of chapter 1 to universes of sets. This is a basic step in order to set up nonstandard measure theory. We shall consider to types of extensions of universes (systems of sets with certain closure properties). The first extension is called enlargement of universes and is one is an axiomatic enlargement of an universe $U$ and is defined as an embedding

$$U \overset{*}{\to} U' \tag{136}$$

which satisfies some axioms. The second one is a generalization of the ultrafilter construction of chapter 1 from the special set $\mathbb{R}$ to universes. It is clear that the construction of a nonprincipal proper ultrafilter of chapter 1 can be extended to any index set $I$. So we seek for some object of the form

$$\text{"}U^* := U^I/\mathcal{F}\text{"}. \tag{137}$$

But before we step into that subject we consider first a more fundamental question. In the preceding chapter we formulated first order logic. and proved the theorem of Loos in the case of real and hyperreal numbers. However if we look at propositions of the working mathematicians we see that quantifiers are applied to sets in general which is covered only by higher order logic. Moreover, it is well-known that elementary structures of mathematics like the natural number system $\mathbb{N}$ cannot be characterized up to isomorphism by a first order axiomatic system. The Peano axiomatic system of the natural numbers is characterized therefore in second order logic.

$$\begin{aligned} &\text{P1} \quad \forall x \neg \sigma(x) = 0 \\[2mm] &\text{P2} \quad \forall x \forall y \, (\sigma(x) = \sigma(y) \to x = y) \\[2mm] &\text{P3} \quad \forall X \, (X(0) \land \forall x \, (X(x) \to X(\sigma(x))) \to \forall X(y)) \end{aligned} \tag{138}$$

(Here sets $X$ are identified with univariate relations).

The axioms (P1)-(P3) characterize the standard arithmetic structure $(\mathbb{N}, \cdot, +, 0, 1)$ up to isomorphism, and it is a result by elementary model theory that this structure cannot be formalized by a finite first order axiomatic system. However, in this section we shall show that nevertheless mathematics can be essentially expressed as a first order theory within a set theoretic axiomatic system expressed by a first order language with one binary relation symbol $\in$. The latter proposition is empirical of course inasmuch it refers to mathematics as a subject of propositions expressed by past present and future mathematicians.

We exemplify it (and make it plausible) by formalization of a theorem by Dedekind

**Theorem 11.** *Two Peano structures are isomorphic.*

which cannot prima facie formalized on a first order level, since

**Theorem 12.** *No model $\mathcal{A}$ with infinite set $A$ is characterizable in first order logic up to isomorphism (i.e. $\Delta$-elementary).*

Proof. Let $\mathcal{A}$ be a model with infinite set $A$ and let $\Phi$ be a set of sentences interpretable in $\mathcal{A}$. Consider the set of models

$$\{\mathcal{B}|\mathcal{B} \cong \mathcal{A}\} \tag{139}$$

The set $\Phi$ has an infinite model. Next recall

**Theorem 13.** *(Löwenheim & Skolem) Let $\Phi$ be a set of expressions which can be satisfied by an infinite model. For any set $A$ there is a model of $\Phi$ which has at least as much elements as $A$ (in the sense that there is an infinite model of $\Phi$ with set $B$ such that there is an injective function from $A$ to $B$).*

Theefore $\Phi$ has a model $\mathcal{B}$ with set $B$ which has the cardinality of he power set of $A$. Since $\mathrm{card}(A) < \mathrm{card}(P(A))$ $\mathcal{B}$ is then not isomorphic to $cal A$. Therefore, Peano structures cannot be characterized by first order theories. Nevertheless we can avoid this consequence somehow. We shall show: in a formal first order language with the set of symbols $S = (U, M, \in)$ with the univariate relation symbols $U$ and $M$ with the interpretation ". is urelement" and ". is set" respectively, and the relation symbol $x \in y$ ("$x$ is element of $y$") it is possible

i) to formalize the theorem of Dedekind in the first order language $S$

ii) to prove the theorem of Dedekind in a first order system.

These facts (and similar experiences) let us believe that first order logic is sufficient in order to express mathematics.

Next we shall make this more precise (plausible) by formulating an axiomatic system in the language $L_S$.

$$\begin{array}{l}
(A0) \quad \forall x (Ux \vee Mx) \\
\text{"every object is either urelement or set"} \\[2mm]
(A1) \quad \forall x \neg (Ux \wedge Mx) \\
\text{"no object is urelement and set at the same time"} \\[2mm]
(A2) \quad \forall x \forall \left( ((Mx \wedge M \wedge \forall z \, (z \in x \leftrightarrow z \in y)) \rightarrow x = y \right) \\
\text{"two sets are equal if the contain the same elements"} \\[2mm]
(A3) \quad \forall x \forall y \exists z \, (Mz \wedge \forall u \, (u \in z \leftrightarrow (u = x \vee u =))) \\
\text{"two sets are equal if the contain the same elements"}
\end{array} \tag{140}$$

In order to make further formalization more readable we introduce the abbreviations

$(\subseteq)$ $\;x \subseteq y \equiv Mx \wedge My \wedge \forall z(z \in x \rightarrow z \in y)$
"$x$ is subset of $y$ "

$(GPzxy)$ $\;GPzxy \equiv Mz \wedge \forall u(u \in z \leftrightarrow (Mu \wedge (\forall v(v \in u \leftrightarrow v = x)$
$\qquad\qquad\qquad \vee \forall v\,(v \in u \leftrightarrow (v = x \vee v = y))))$
"$z$ is the ordered pair of $x$ and $y$"

$(GTuxyz)$ $\;\forall x \forall\,(((Mx \wedge M \wedge \forall z\,(z \in x \leftrightarrow z \in y)) \rightarrow x = y)$
"$u = (x, y, z)$"

$(GPEuxy)$ $\;\forall x \forall y \exists z\,(Mz \wedge \forall u\,(u \in z \leftrightarrow (u = x \vee u =)))$
"$(x, y) \in u$"

$(Fu)Fu \equiv Mu \wedge \forall z(z \in u \rightarrow \exists x \exists y GPzxy) \wedge \forall x \forall y \forall y'\,((Euxy \wedge Euxy') \rightarrow y = y'))$
"$u$ is function

$\mathrm{Df}v \equiv Ff \wedge Mv \wedge \forall x\,(x \in v \leftrightarrow \exists y Efxy)$
"$v$ is domain of the function $uf$"

$(Buv)Buv \equiv Fu \wedge Mv \wedge \forall y(y \in v \leftrightarrow \exists x Euxy)$
"$v$ is image of the function $u$."

$$(141)$$

Next we use the preceding abbreviations in order to formalize Peano structures.

# 14 Universes and enlargements

The axiomatic system RZFC tells us what sets are. In set theory sometimes you take entities for granted (as Kronecker expresses that the natural numbers are given by god, and the rest is constructed by mankind). Similar in nonstandard theory we tend to consider some set as the set of urelements. In ZFC people realised that urelements are superfluous. However, when we talk about universes over ..., we include urelements as convenient. We denote sets by capital letters and elements which are either sets or urelements by small letters. E.g. in $A \in \mathbb{U}$ $A$ is a set, while in $a \in \mathbb{U}$ $a$ may be a set or an urelement. A universe is a set with certain properties. First if $A \in \mathbb{U}$ is a set, then we want all elements of $a$ to be present in $\mathbb{U}$, i.e.

$$a \in A \in \mathbb{U} \quad \Rightarrow \quad a \in \mathbb{U}. \qquad (142)$$

Any set $\mathbb{U}$ which satisfies 142 is called transitive. Furthermore, in a universe we want to have with a set $A \in \mathbb{U}$ its transitive closure $\mathrm{Tr}(A) \in \mathbb{U}$. Here $\mathrm{Tr}(A)$ is the smallest transitive set that contains $A$. If $A$ is transitive itself, then $A = \mathrm{Tr}(A)$, of course. We require:

if $A \in \mathbb{U}$, then there ex. a transitive set $B \in \mathbb{U}$ with $A \subset B \subset \mathbb{U}$. $\qquad (143)$

Finally we require (and are allowed to by RZFC) that

$$\text{if } a, b \in U, \text{ then } \{a, b\} \in \mathbb{U}$$

$$\text{if } A, B \in \mathbb{U} \text{ are sets, then } A \cup B \in \mathbb{U} \tag{144}$$

$$\text{if } A \in \mathbb{U} \text{ is a set, then } P(A) \in \mathbb{U}$$

If the universe contains a set $\mathbb{S}$ such that the members of $S$ are individuals in the sense that

$$\forall x \in \mathbb{S} \left[ x \neq \oslash \wedge (\forall y \in \mathbb{U}(y \notin x)) \right]. \tag{145}$$

The next step of course is to show that universes exist. They are realized by superstructures. Let $\mathbb{S}$ be a set. We define a series cumulative power set by

$$\mathbb{U}_0 (\mathbb{S}) = \mathbb{S},$$

$$\mathbb{U}_{n+1} (\mathbb{S}) = \mathbb{U}_n (\mathbb{S}) \cup P (\mathbb{U}_n (\mathbb{S})). \tag{146}$$

Then it is easy to check that

$$\mathbb{U} (\mathbb{S}) := \bigcup_{n \in \mathbb{N}} \mathbb{U}_n (\mathbb{S}) \tag{147}$$

is a universe. If the set $\mathbb{S}$ is known from the context, then we . The language $\mathcal{L}_R$ with quantification restricted to the sets of the universe $\mathbb{U}$ is denoted $\mathcal{L}_R^{\mathbb{U}}$. Next, a nonstandard framework for a set $\mathbb{S}$ comprises a universe $\mathbb{U}$ over $\mathbb{S}$ and a map

$$\mathbb{U} \overset{*}{\to} \mathbb{U}', \tag{148}$$

which satisfies

$$a^* = a \text{ for } a \in \mathbb{S},$$

$$\oslash^* = \oslash, \tag{149}$$

$$\text{the } \mathcal{L}_R^{\mathbb{U}}\text{-sentence } \phi \text{ is true iff } \phi^* \text{ is true.}$$

Such a nonstandard framework is called an enlargement if the following condition is satisfied:

if $A \in \mathbb{U}$ is a collection of sets with the finite intersection property, then there exists an element $z \in \mathbb{U}'$ such that

$$z \in \bigcap \{Z^* | Z \in A\}. \tag{150}$$

**Example 14.** *If $\mathbb{U}$ is a universe on $\mathbb{R}$ and $A$ is the set of intervals*

$$\{(0, r) | r > 0\} \tag{151}$$

*then the latter set satisfies the finite intersection property. Then the enlargement principle tells us that there exists a positive infinitesimals, i.e. there exists*

$$b \in \cap \{(0, r)^* | r > 0\} = \text{'set of positive infinitesimals'}. \tag{152}$$

**Example 15.** *Consider a universe $U$ on $\mathbb{N}$ and an enlargement $U \xrightarrow{*} U'$. The set*

$$A = \{\mathbb{N}_{\geq n} | n \in \mathbb{N}\} \tag{153}$$

*with $\mathbb{N}_{\geq n} := \{m \in \mathbb{N} | m \geq n\}$ satisfies the finite intersection property. Then the enlargement principle tells us that*

$$\exists b : b \in \cap\{\mathbb{N}_{\geq n}^* | n \in \mathbb{N}\} = \mathbb{N}^* \setminus \mathbb{N}, \tag{154}$$

*i.e. there is a set of unlimited numbers.*

The question now is whether enlargements really exist. This can be shown with the ultrafilter construction. Two types of sets are of special interest for us: the first type are the internal sets:

$$a \in \mathbb{U} \text{ is internal if } a \in A^* \text{ for some } A \in \mathbb{U}; \tag{155}$$

the second type are the hyperfinite sets: let

$$P_F(A) = \{B \subseteq A | B \text{ is finite }\}. \tag{156}$$

Then the hyperfinite sets are the members of $P_F(A)^* \in \mathbb{U}'$.

# 15    The ultrafilter construction of enlargements

Let $I$ be an infinite set and let $\mathcal{F}$ be a nonprincipal ultrafilter on $I$. Let $\mathbb{S}$ be a set and let let $\mathbb{U}$ be a universe over $\mathbb{S}$. To $a \in \mathbb{U}$ assign $a_I \in \mathbb{U}^I$, the function with constant value $a$. This way we embed $\mathbb{U}$ in a larger universe. Similarly as in the ultrafilter construction on $\mathbb{R}$ we have to consider equivalence of elements with respect to the ultrafilter, this time of functions $f, g \in \mathbb{U}^I$. We say

$$f \sim g \text{ iff } \{i | f(i) = g(i)\} \in \mathcal{F}$$
$$f \in g \sim f' \in g' \text{ iff } \{i | f(i) \in g(i) \& f'(i) \in g'(i)\} \in \mathcal{F} \tag{157}$$

We denote equivalence classes b [.] as before and define

$$\mathcal{W}_n := \left\{ f \in \mathbb{U}_n^I | \{i | f(i) \in \mathbb{U}_n\} \in \mathcal{F} \right\}. \tag{158}$$

$$\mathcal{W} := \cup_{n \in \mathbb{N}} \mathcal{W}_n \tag{159}$$

Now for $f \in \mathcal{W}_0$ let $[f] := \{[h] | h \in \mathcal{W}_0|\}$, and let

$$\mathbb{Y} = \{[f] | f \in \mathcal{W}_0\} \tag{160}$$

This defines

$$\mathbb{U}_0(\mathbb{Y}) = \mathbb{Y} \tag{161}$$

Inductively, having defined $\mathbb{U}_n(\mathbb{Y})$, for $f \in \mathcal{W}_{n+1} \setminus \mathcal{W}_n$ define

$$[f] = \{[h] | h \in \mathcal{W}_n \text{ and } \{i | h(i) \in f(i)\} \in \mathcal{F}\} \tag{162}$$

Then

$$\mathbb{U}_{n+1}(\mathbb{Y}) = \mathbb{U}_n(\mathbb{Y}) \cup \{[f] | f \in \mathcal{W}_{n+1} \setminus \mathcal{W}_n\}, \tag{163}$$

and
$$\mathbb{U}(\mathbb{Y}) = \cup_{n \in \mathbb{N}} \mathbb{U}_n(\mathbb{Y}). \tag{164}$$

Now, $\mathbb{U}(\mathbb{Y})$ is the ultrafilter-enlargement we had looked for. For each $f, g \in \mathcal{W}$ itis easy to see that

$$[f] \in [g] \text{ iff } \{i | f(i) \in g(i)\} \in \mathcal{F}, \ [f] = [g] \text{ iff } \{i | f(i) = g(i)\} \in \mathcal{F} \tag{165}$$

The map
$$* : \mathbb{U}(\mathbb{X}) \to \mathbb{U}(\mathbb{Y}), \ a \to a^* = [a_I] \tag{166}$$

is an embedding of the universe $\mathbb{U}(\mathbb{X})$ in the universe $\mathbb{U}(\mathbb{Y})$, and we observe that

$$\oslash^* = \oslash, \text{ and } \mathbb{X}^* = \mathbb{Y}. \tag{167}$$

Since $\mathcal{F}$ is an ultrafilter, we know that the enlargement $\mathbb{U}(\mathbb{Y})$ has nonstandard members. Let $L_{\mathbb{U}(\mathbb{X})}$ and $L_{\mathbb{U}(\mathbb{Y})}$ be the formal languages of the respective universes. Denote the model of the ultrafilter-enlargement by $\mathcal{U}_{\mathbb{Y}} = (\mathbb{U}(\mathbb{Y}), \in)$ the model of the original universe by $\mathcal{U}_{\mathbb{X}} = (\mathbb{U}(\mathbb{X}), \in)$. Then we get the following version of the theorem of Loos.

**Theorem 16.** *For any $L_{\mathbb{U}(\mathbb{X})}$-formula $\phi(x_1, \cdots, x_m)$ and $f_1, \cdots f_m \in \mathcal{W}$*

$$\mathcal{U}_{\mathbb{Y}} \models \phi\left(e|_{[f_1], \cdots, [f_m]}^{x_1, \cdots, x_m}\right) \textit{ iff } \left\{i | \mathcal{U}_{\mathbb{X}} \models \phi\left(e^i|_{f_1(i), \cdots, f_m(i)}^{x_1, \cdots, x_m}\right)\right\} \in \mathcal{F} \tag{168}$$

Let $U$ be a universe (which contains the real numbers as individuals) and let $U \xrightarrow{*} U'$ be an enlargement. For $A \in U$ and let

$$P_F(A) = \{b \subseteq A | B \text{ is finite }\}. \tag{169}$$

$P_F(A)^*$ are called hyperfinite subsets of $A$. As an example, consider $P_F(\mathbb{N}$ and the $L_U$-sentence

$$\forall n \in \mathbb{N} \exists A \in P_F(\mathbb{N}) \forall m \in \mathbb{N} [m \in A \leftrightarrow m \leq n], \tag{170}$$

i.e. the sentence which has the meaning that for each natural number $n \in \mathbb{N}$ there is a set $A = \{1, \cdots, n\}$ in $P_F(\mathbb{N})$. The transfer sentence is

$$\forall n \in \mathbb{N}^* \exists A \in P_F(\mathbb{N})^* \forall m \in \mathbb{N}^{(} [m \in A \leftrightarrow m \leq n]. \tag{171}$$

Hence, for all $n \in \mathbb{N}^*$
$$A = \{1, \cdots, n\} \in P_F(\mathbb{N})^*. \tag{172}$$

Note that $n$ can be infinite, i.e. $n \in \mathbb{N}^* \setminus \mathbb{N}$. We prove

**Theorem 17.** *$A$ is hyperfinite iff there exists $n \in \mathbb{N}^*$ and an internal bijection*

$$f : \{1, \cdots, n\} \to A. \tag{173}$$

Here a function $f : A \to B$ is called internal if the set graph$(f) \subseteq A \times B$ is internal.

Proof. Consider a $L_{\mathbb{U}}$-formula

$$\phi(X, Y, n, f) \tag{174}$$

31

which expresses that $f : X \to Y$ with $X = \{m \in \mathbb{N} | m \leq n\}$ is a bijection. Then the $L_{\mathbb{U}}$-sentence

$$\psi \equiv \forall Y \in P_F(B) \exists n \in \mathbb{N} \exists f \in P(\mathbb{N} \times B) \exists X \in P(\mathbb{N}) \phi(X, Y, n, f) \qquad (175)$$

asserts that for all $Y \in P_F(B)$ there is a number $n$ and a bijection between $X = \{1, \cdots, n\}$ and $Y$- a sentence which is true. The sentence $\psi^*$ is true by transfer. So if $B \in \mathbb{U}$ and $A \in P_F(B)^*$ then the claim follows from the truth of $\psi^*$. For the converse suppose that there is an internal bijection $f : X = \{1, \cdots n\} \to A$ for some $n \in \mathbb{N}^*$. Then $A$ is internal, because it is the range of an internal function. We want to show that $A$ is hyperfinite. First we observe that

$$\exists X \in P(\mathbb{N})^* \phi(X, A, n, f) \qquad (176)$$

is true. Hence the claim that $A$ is hyperfinite follows from transfer of the true $L_{\mathbb{U}}$-sentence

$$\forall Y \in B \exists n \in \mathbb{N} \exists f \in P(\mathbb{N} \times A) \exists X \in P(\mathbb{N}) (\phi(X, A, n, f) \to Y \in P_F(A)). \qquad (177)$$

# 16 Nonstandard proof of the central limit theorem

**Theorem 18.** *Let the random variables $X_n$ be binomially distributed with parameters $p$ and $q = 1 - p$. If for $n \in \mathbb{N}^* \setminus \mathbb{N}$ is $m \in \mathbb{N}^*$ is such that*

$$x = \frac{m - np}{\sqrt{npq}} \qquad (178)$$

*is in a standard interval $[a, b]$ , then*

$$\sqrt{npq} B_n(m) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2}. \qquad (179)$$

Proof. Consider the classical formula of Stirling for integer $s \in \mathbb{N}$

$$s! = \sqrt{2\pi s} s^s e^{-s} \exp(\Theta_s), \qquad (180)$$

where $|\Theta_s| \leq \frac{1}{12s}$. For hyperfinite $s \in \mathbb{N}^* \setminus \mathbb{N}$ the formula becomes

$$s! \approx \sqrt{2\pi s} s^s e^{-s}. \qquad (181)$$

Since $x$ is bounded and $m = np + x\sqrt{npq}$, and $n - m = nq + x\sqrt{npq}$, we have $\frac{1}{m} \approx 0$ and $\frac{1}{n-m} \approx 0$ for $n, m \in \mathbb{N}^* \setminus \mathbb{N}$. For such $n, m$ (181) gives

$$B_n(m) = \frac{n!}{m!(n-m)!} p^m q^{n-m} = \sqrt{\frac{n}{2\pi m(n-m)}} \frac{n^n}{m^m} \frac{p^n q^{n-m}}{m^m (n-m)^{n-m}} \qquad (182)$$

Now, consider

$$\ln \frac{n^n}{m^m} \frac{p^n q^{n-m}}{m^m (n-m)^{n-m}} = \ln \left(\frac{np}{m}\right)^m + \ln \left(\frac{nq}{n-m}\right)^{n-m}$$

$$- \left(np + x\sqrt{npq}\right) \ln \left(1 + x\sqrt{\frac{q}{np}}\right) - \left(nx - x\sqrt{npq}\right) \ln \left(1 - x\sqrt{\frac{q}{np}}\right) \qquad (183)$$

$$\approx -\sqrt{npq} x - x^2 q + x^2 \frac{q}{2} + \sqrt{npq} x - x^2 p + x^2 \frac{p}{2} = -\frac{x^2}{2}.$$

# 17 General Topologies

First we define general topologies on $X \in U$ where $U$ is some standard universe in the enlargement $U \xrightarrow{*} U'$

**Definition 19.** *Let $\oslash \neq X \in U$. A base at $x \in X$ is a set $\mathcal{B}_x \in P(X)$ is such that*

$$\forall U, V \in \mathcal{B}_x \exists W \in \mathcal{B}_x : x \in W \subseteq U \cap V \tag{184}$$

**Definition 20.** *For $x \in X$ a monad is a set*

$$md(x) := \cap_{U \in \mathcal{B}_x} U^* \tag{185}$$

*We write $y \sim x$ if $y \in md(x)$.*

**Definition 21.** *$O \in X$ is called open, if*

$$\forall x \in O \exists U \in \mathcal{B}_x : U \subseteq O, \tag{186}$$

*or, equivalently,*

$$\forall x \in O \ md(x) \subseteq O^*. \tag{187}$$

To see the equivalence, first observe that assuming (186) for $x \in O$ there is $U \in \mathcal{B}_x$ such that $U \subseteq O$. Hence $md(x) \subseteq U^* \subseteq O^*$. On the other hand, assuming (187), if $md(x) \subseteq O^*$ then $\exists W \in \mathcal{B}_x^*$ with $W \subset md(x) \subseteq O^*$ by the definition of monad above and the finite intersection property of $\mathcal{B}_x^*$. Hence the judgement

$$\forall x \in O^* \exists W \in \mathcal{B}_x^* : W \subseteq md(x) \ \& \ md(x) \subseteq O^* \tag{188}$$

is true as is the logical consequence

$$\forall x \in O^* \exists W \in \mathcal{B}_x^* : W \subseteq O^*. \tag{189}$$

Hence,

$$\forall x \in O \exists W \in \mathcal{B}_x : W \subseteq O \tag{190}$$

holds by transfer. The collection of all open sets $\mathcal{T}$ is called topology. It is an immediate consequence of (187) that the empty set $\oslash$ and the set $X$ are open sets Furthermore arbitrary unions of open sets are open. Next

**Definition 22.** *Given $\mathcal{B}_x$ for each $x \in X$ and $A \subseteq X$ we say that $x$ belongs to the closure of $A$ (in symbols $\overline{A}$) if*

$$md(x) \cap A^* \neq \oslash. \tag{191}$$

*Hence, we define*

$$\overline{A} := \{x \in X | md(x) \cap A^* \neq \oslash\} \tag{192}$$

It is easily verified that the arbitrary intersection of claosed sets is closed. Note that $\oslash$ and $X$ are also closed; sometimes they are called clopen (i.e. open and closed). Now, we can define continuous maps for general topologies.

**Definition 23.** *If $(X, \mathcal{T}), (Y, \mathcal{S})$ are topological spaces, then we call $f : \mathcal{T} \to (Y, \mathcal{S})$ continuous if*

$$\forall V \in \mathcal{T} \left( f(x) \in V \to \exists U \in \mathcal{S} : x \in U \& f(U) \subseteq V \right) \tag{193}$$

*or equivalently and more succinct,*

$$f\left(md_{\mathcal{S}}(x)\right) \subseteq md_{\mathcal{T}}\left(f(x)\right). \tag{194}$$

As an example of a topology which is not induced by a metric, consider the set

$$X = \{f : [0,1] \to [0,1]\} \tag{195}$$

and for $f \in X$

$$B_f := \left\{ U_{\epsilon, x_1, \cdots, x_n} | \epsilon > 0, n \in \mathbb{N}, x_1, \cdots, x_n \in [0,1] \right\}, \tag{196}$$

where

$$U_{\epsilon, x_1, \cdots, x_n} := \{g \in X | |g(x_i) - g(y_i)| < \epsilon, i \in \mathbb{N}_n\} \tag{197}$$

So far we have used nonstandard extensions $X^*$ of a topological spaces $X$ in order to define topologies on $X$. The topology on $X^*$ which is used most in the literature is the so-called $S$-topology, which has the base $\mathcal{B} = \{U^* | U \in \mathcal{T}\}$. Note that this topology is not Hausdorff since it cannot separate two different points of a monad. Next recall that a set $A \subseteq X$ is compact if every open cover of $A$ has a finite subcover. Robinson observed the following nonstandard criterion for compactness, the proof of which is left to the reader.

**Theorem 24.** *A set $A \subseteq X$ is compact iff for each $y \in A^*$ there is an $x$ with $y \in md(x)$.*

# 18 Introduction to measure theory

Measures assign values to sets, but not to every set: this is the price of our generosity in RZFC with respect to sets. The collection of sets measures are defined on are usually $\sigma-$algebras. Nonstandard measures are defined on collections of internal sets. However, there is one problem: countable unions of internal sets are usually not internal sets. E.g. each singleton $\{n\}$, where $n \in \mathbb{N}$ is a is an internal set but $\mathbb{N} = \cup_{n \in \mathbb{N}} \{n\}$ is not an internal set, whenever $\mathbb{N}^* \setminus \mathbb{N} \neq \oslash$. For this reason, the countable union of members of an internal $\sigma$-algebra can only belong to the $\sigma$-algebra if it equals the union of finitely many of its terms. This was the essential observation of Loeb. We next review standard measure theory

# 19 Rings, Algebras, and Measures

Let $S$ be a set. A ring of sets is a nonempty collection $\mathcal{A}$ of subsets of a set $S$ that is closed under set differences and unions, i.e.

$$\text{if } A, B \in \mathcal{A} \text{ then } A \setminus B, A \cup B \in \mathcal{A}. \tag{198}$$

Note that this implies that $\oslash \in \mathcal{A}$ and that $\mathcal{A}$ is closed under symmetric differences and intersections. An algebra is a ring with $S \in \mathcal{A}$. A $\sigma$-ring is a ring with closure under countab;e union, i.e.

$$A_n \in \mathcal{A} \text{ for all } n \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}. \tag{199}$$

**Example 25.** *Let $N \in \mathbb{N}^* \setminus \mathbb{N} \neq \oslash$, $S := \{1, \cdots, N\}$. $P_i(S) := \{A \subseteq S | A \text{ is internal } \}$ is an algebra, but not a $\sigma$-algebra: we have $A_n := \{1, \cdots, n\} \in P_i(S)$, but $\cup_{n \in \mathbb{N}} A_n = \mathbb{N}$ is external.*

**Example 26.** *Recall that a countably saturated enlargement* $\mathbb{U} \xrightarrow{*} \mathbb{U}'$ *is an enlargement, where for each sequence* $(A_n)_{n \in \mathbb{N}}$ *of internal sets with* $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_i \supseteq A_{i+1} \supseteq \cdots$ *we have* $\bigcap_{i \in \mathbb{N}} A_i \neq \oslash$. *If* $\mathcal{A} \in \mathbb{U}$ *is an algebra, then* $\mathcal{A}^* \in \mathbb{U}'$ *is an algebra by transfer, but not a* $\sigma$*-algebra.*

Measures assign values in $\overline{\mathbb{R}} := \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$.

# 20 The Loeb measure

Let $(\Omega, \mathcal{A}, P)$ be an internal, finitely additive probability space, i.e.

i) $\Omega$ internal

ii) $\mathcal{A}$ is internal subalgebra of $\mathcal{P}(\Omega)$

iii) $P : \mathcal{A} \to {}^*\mathbb{R}$ is an internal function such that

iv) $P(\oslash) = 0$, $P(\Omega) = 1$, $\forall A, B$ : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

The following theorem is the main theorem of non standard probability theory.

**Theorem 27.** *There is a standard ($\sigma$-additive) probability space* $(\Omega, \mathcal{A}_L, P_L)$ *such that*

*i)* $\mathcal{A}_L$ *is a* $\sigma$*-algebra with* $\mathcal{A} \subseteq \mathcal{A}_L \subseteq \mathcal{P}(\Omega)$

*ii)* $P_L = P^\circ$ *on* $\mathcal{A}$

*iii) For every* $A \in \mathcal{A}_L$ *and standard* $\epsilon > 0$ *there are* $\mathcal{A}_i$ *and* $\mathcal{A}_o$ *in* $\mathcal{A}$ *such that* $A_i \subseteq A \subseteq A_o$ *and* $P(A_o \setminus A_i) < \epsilon$

*iv) For every* $A \in \mathcal{A}_L$ *there is* $B \in \mathcal{A}$ *such that* $P_L(A \Delta B) = 0$

*The space* $(\Omega, \mathcal{A}_L, P_L)$ *is called a Loeb probability space*

Proof. We start with the finitely additive function $P : \mathcal{A} \to [0, \infty]^*$, where $\mathcal{A}$ is an internal ring of subsets of the internal set $S$. Then we set for all $A \in \mathcal{A}$

$$P_L(A) = \begin{cases} \text{sh}(P(A)), & \text{if } P_L(A) \text{ is limited} \\ \infty & \text{else.} \end{cases} \tag{200}$$

We extend $\mathcal{A}$ to a $\sigma$-algebra by the standard outer measure construction. If $B \subset S$ is an arbitrary subset of $S$, then define

$$P_L(B) = \inf \left\{ \sum_{n \in \mathbb{N}} P(A_n) | A_n \in \mathcal{A} \text{ and } B \subseteq \cup_{n \in \mathbb{N}} A_n \right\} \tag{201}$$

A set $B \subset S$ is called Loeb-measurable ($P_L$-measurable if it splits every set $E \subset S$ $P_L$-additively in the sense that

$$P_L(E) = P_L(E \cap B) + P_L(E \setminus B). \tag{202}$$

We define

$$\mathcal{A}_L := \{A \subseteq S | A \text{ is } P_L - \text{measurable}\} \tag{203}$$

From the standard outer measure construction we know that $\mathcal{A}_L$ is a $\sigma$-algebra and $P_L$ is a complete measure. We say that an arbitrary set $B \subseteq S$ is called $P$-approximable if for every $\epsilon > 0$ there exist approximating sets $C_\epsilon, D_\epsilon \in \mathcal{A}$, such that

$$C_\epsilon \subseteq B \subseteq D_\epsilon \text{ and } P(D_\epsilon \setminus C_\epsilon) < \epsilon \tag{204}$$

First we show

**Lemma 28.** *If $B$ is $P$-approximable then there exists a set $A \in \mathcal{A}$ such that*

$$P_L(A \Delta B) = 0. \tag{205}$$

We apply $P$-approximability and find two sequences of elements of $\mathcal{A}$ $(C_i)_{i \in \mathbb{N}}$ and $(D_i)_{i \in \mathbb{N}}$ with

$$\cdots \subseteq C_n \subseteq C_{n+1} \subseteq \cdots \subseteq B \subseteq \cdots \subseteq D_n \subseteq D_{n-1} \tag{206}$$

and

$$P_L(D_n \setminus A_n) < \frac{1}{n} \tag{207}$$

Recall that the embedding $U \overset{*}{\to} U'$ is sequentially comprehensive which means that any function $f : \mathbb{N} \to B \in U^*$ extends to an internal function $f : \mathbb{N}^* \to B \in U^*$. For each $k \in \mathbb{N}$ we have

$$\forall n \in \mathbb{N}^* \, (n \le k \to C_n \subseteq D_k \subseteq D_n). \tag{208}$$

Since all constants are internal the latter statement is internal. Hence there exists $K \in \mathbb{N}^* \setminus \mathbb{N}$ such that

$$\forall n \in \mathcal{N} : C_n \subseteq C_K \subseteq D_n \tag{209}$$

holds. But then for all $n \in \mathcal{N}$ we have

$$D_K \Delta B = (D_K \setminus B) \cup (B \setminus D_K) \subseteq D_n \setminus C_n \tag{210}$$

and

$$P_L(D_K \Delta B) < \frac{1}{n}. \tag{211}$$

Hence $P_L(D_K \Delta B) = 0$ while $D_K \in \mathcal{A}$. Next we show: if $B$ is Loeb measurable, then $B$ is $P$-approximable. This follows from observation that for each Loeb-measurable set $B$ we have

$$
\begin{aligned}
P_L(B) \ &= \inf\{P_L(A) | B \subseteq A \in \mathcal{A}\} \\
&= \sup\{P_L(A) | A \subseteq B \& A \in \mathcal{A}\}.
\end{aligned}
\tag{212}
$$

If (212) holds, then given $\epsilon \in \mathbb{R}_+$ there are $C_\epsilon, D_\epsilon \in \mathcal{A}$ s. th.

$$C_\epsilon \subseteq B \subseteq D_\epsilon \,, P_L(D_\epsilon < P_L(B) + \frac{\epsilon}{2}, \text{ and } P_L(C_\epsilon) < P_L(B) + \frac{\epsilon}{2}. \tag{213}$$

Then we get

$$
\begin{aligned}
P_L(D_\epsilon \setminus C_\epsilon) \ &= P_L(D_\epsilon \setminus B \cup B \setminus C_\epsilon) \\
&= P_L(D_\epsilon \setminus B) + P_L(B \setminus C_\epsilon) \\
&= P_L(D_\epsilon) - P_L(B) + P_L(B) - P_L(C_\epsilon) \\
&< \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon,
\end{aligned}
\tag{214}
$$

as desired. In order to prove (212) we first show that

$$\forall \epsilon \in \mathbb{R}_+ \exists A_\epsilon \in \mathcal{A} : B \subseteq A_\epsilon \ \& \ P_L(A_\epsilon) \leq P_L(B) + \epsilon. \tag{215}$$

Now, by the properties of outer measure for $\epsilon > 0$ there exist a nondecreasing sequence $(A_i)_{i \in \mathbb{N}}$ of elements of A with

$$\cup_{i \in \mathbb{N}} A_i \supseteq B \text{ and } P_L \left( \cup_{i \in \mathbb{N}} A_i \right) < P_L(B) + \epsilon. \tag{216}$$

By sequential comprehensiveness, the sequence $(A_i)_{i \in \mathbb{N}}$ extends to a sequence $(A_i)_{i \in \mathbb{N}^*}$ of elements in $\mathcal{A}$. For each $k \in \mathbb{N}$ we have

$$kk \forall n \in \mathcal{N}^* \left( n \leq k \rightarrow A_n \subseteq A_k \ \& \ P(A_n) < P_L(B) + \epsilon \right), \tag{217}$$

where we observed that $P(A_n) \sim P_L(A_n) \leq P_L(\cup_{i \in \mathbb{N}})$. The statement (**??**) is internal, so by overflow it must be true for some $K \in \mathbb{N}^* \setminus \mathbb{N}$ in place of $k \in \mathbb{N}$. Hence, $A_n \subseteq A_K$ for all $n \in \mathcal{N}$ and $B \subseteq \cup_{n \in \mathbb{N}} A_n \subseteq A_K$, while $P_L(A_K) \sim P(A_K) < P_L(B) + \epsilon$, so $A_K$ is the set $A_\epsilon$ we are looking for. Next, to show that

$$P_L(B) = \sup\{P_L(A) | A \subseteq B \& A \in \mathcal{A}\} \tag{218}$$

means to show that given $\epsilon \in \mathbb{R}_+$ there is $A_\epsilon \subseteq B$ with $P_L(B) - \epsilon < P_L(A_\epsilon)$. By the argument above there is $D \in \mathcal{A}$ with $B \subseteq D$. Since $D, B$ are Loeb-measurable $D \setminus B$ is Loeb-measurable, so by the argument above there is $C \in \mathcal{A}$ with $C \supseteq D \setminus B$ s.th.

$$P_L(C) < P_L(D \setminus B) + \epsilon. \tag{219}$$

W.l.o.g. $C \subseteq D$ (otherwise replace $C$ by $C \cap D$). Let $A_\epsilon = D \setminus C \in \mathcal{A}$. Then $A_\epsilon \subseteq B$ and $C$ is the disjoint union of $D \setminus B$ and $B \setminus A_\epsilon$. Hence,

$$P_L(D \setminus B) + P_L(B \setminus A_\epsilon) = P_L(C) < P_L(B \setminus A_\epsilon) < \epsilon. \tag{220}$$

Hence,

$$P_L(B) = P_L(A_\epsilon) + P_L(B \setminus A_\epsilon) < P_L(A_\epsilon) + \epsilon, \tag{221}$$

so $P_L(B) - \epsilon < P_L(A_\epsilon)$, as desired. Next we prove: if $B$ is P-approximable, then $B$ is Loeb-measurable. We have to show that for $P$-approximable $B$ and for all $A \subseteq S$

$$P_L(E) \geq P_L(E \cap B) + P_L(E \setminus B). \tag{222}$$

By Lemma (28) we have

$$\exists A \in \mathcal{A} : P_L(A \Delta B) = 0. \tag{223}$$

So

$$P_L(E) \geq P_L(E \cap A) + P_L(E \setminus A). \tag{224}$$

Let $C := (E \cap B) \setminus A$, $D = (E \cap A) \setminus B$, $G = E \setminus (A \cup B)$, and $H = E \cap A \cap B$. Then $C \subseteq B \setminus A$ and $D \subseteq A \setminus B$. Hence, $P_L(C) = P_L(D) = 0$. Furthermore, $C \cup H = E \cap B$, while $C \cap H = \oslash$. Hence,

$$P_L(E \cap B) = P_L(C \cup H) \leq P_L(C) + P_L(H) = P_L(H), \tag{225}$$

and

$$P_L(E \setminus B) = P_L(D \cup G) \leq P_L(D) + P_L(G) = P_L(G). \tag{226}$$

Hence,

$$
\begin{aligned}
P_L(E \setminus B) + P_L(E \setminus B) \ &\leq P_L(H) + P_L(G) \\
&\leq P_L(E \cap A) + P_L(E \setminus A) \qquad (227) \\
&= P_L(E). \ \square
\end{aligned}
$$

## 21  Stochastic Integrals, Itô formulas

Definition of stochastic integrals and proofs of Itô formulas are particularly simple in the nonstandard framework. Recall that an internal process is just an internal map

$$X : \Omega \times T \to^* \mathbb{R}^n, \qquad (228)$$

where $(\Omega, \mathcal{F}, P)$ is an internal probability space. We may introduce stochastic integrals as follows:

**Definition 29.** *Let* $X : \Omega \times T \to^* \mathbb{R}$ *and* $Y : \Omega \times T \to^* \mathbb{R}$ *be two internal processes. Then the internal process*

$$\int Y \, dX := \sum_{s<t} X(s,.) \Delta Y(s,.) \qquad (229)$$

*is called a stochastic integral. Here,* $\Delta Y(t,.) = Y(t,.) - Y(t - \Delta t,.)$.

Note that in nonstandard analysis an alternative way of defining martingales $M = (M_t)_{t \geq 0}$ is to require

$$\int_{[\omega]_t} \Delta M_t(\omega) dP(\omega) = 0 \qquad (230)$$

for all $t \geq 0$ and all $\omega$, and where $M$ is an internal process. Recall that a martingale $M$ is called a $\lambda^2$-martingale if $E\left[M_t^2\right]$ is finite for all $t \in T$. Furthermore recall that an internal process $Y$ is called $S$-continuous, if the internal map $t \to X_t(\omega)$ is $S$-continuous for almost all $\omega$. Here an internal map $f : T \to {}^*\mathbb{R}$ is $S$-continuous if for all $s, t$ with $s \approx t$ (i.e., $s - t$ infinitesimal) we have

$$-\infty <^\circ f(s) =^\circ f(t) < \infty. \qquad (231)$$

Note that in non-standard analysis the quadratic variation of $M$ can be defined by

$$[M]_t(\Omega) = \sum_{s<t} \Delta M_s(\omega)^2. \qquad (232)$$

**Theorem 30.** *(Itô's formula) Let* $M$ *be an internal process, which is a* $S$-*continuous* $\lambda^2$-*martingal, and assume that* $f : \mathbb{R} \to \mathbb{R}$ *is twice continuously differentable. Then*

$$^*f(M_t) \approx^* f(M_0) + \int_0^t {}^*f(M_s) dM_s + \tfrac{1}{2} \int_0^t {}^*f''(M_s) d[M]_s \qquad (233)$$

**Proof.** From the nonstandard Taylor formula we compute the increment

$$
\begin{aligned}
{}^*f(M_s + \Delta M_s) -^* f(M_s) =^* f'(M_s)\Delta M_s + \tfrac{1}{2}{}^* f'' \Delta M_s^2 \\
+ \tfrac{1}{2}\left(f'' \Delta \Xi_s^2 - f'' \Delta M_s^2\right)
\end{aligned}
\qquad (234)
$$

for some $\Xi_s$ between $M_s$ and $M_s + \Delta M_s$. Summing up a telescopic sum leads to the result because the sum

$$\sum_{s<t} \frac{1}{2} \left( f'' \Delta \Xi_s^2 - f'' \Delta M_s^2 \right) \tag{235}$$

is infinitesimal. ∎

## 22   Proof of generalized Feyman-Kac formulas and generalized stochastic processes

In this section we shall prove Feynman-Kac formulas for generalized Brownian motion. We also introduce generalized processes related to parabolic systems with variable coefficients. This is a first step in the direction of a stochastic field theory which may be worked out in the future. We have already seen that generalized Brownian motions can be defined on the same path space $\Omega_n$ as the vectoriell Brownian motion, i.e., on the path space

$$\Omega_n = \{ \omega : T \to \{-1, 1\}^n \,|\, \omega \text{ internal } \}. \tag{236}$$

Recall that we defined an equivalence relation on $\Omega_n$ by

$$\omega \sim_n \tilde{\omega} \text{ iff } \forall s < t : \omega(s) = \tilde{\omega}(s), \tag{237}$$

and denoted the corresponding equivalence classes by $[\omega]_t^n$. Furthermore, on

$$\Omega_M^n = \left\{ [\omega]_{t_M}^n \,|\, \omega \in \Omega_n \right\}, \tag{238}$$

we defined an internal probability measure

$$P_M^n : \Omega_M^n \to [0, 1]$$
$$P_M^n([\omega]_{t_M}^n) = \frac{1}{2^{nM}} \text{ for all } [\omega]_{t_M}^n \in \Omega_M^n. \tag{239}$$

Let us consider the scalar case first, and how Feynman-Kac formulas can be proved in the framework of nonstandard analysis. For each time $t_M$ let us define the random variable

$$B_{t_M}^x : \Omega_{t_M}^n \to^* \mathbb{R}^n$$
$$B_{t_M}^x([\omega]_{t_M}^n) := x + \sum_{i=1}^n \left( \sum_{s<t} \omega_i(s) \sqrt{\Delta t} \right) \mathbf{e}_i. \tag{240}$$

Next recall that $^*\mathbb{R}_{\Delta x}$ denotes a hyperfinite discretization of $^*\mathbb{R}$ with respect to the hyperreal number $\Delta x$. In case of dimension $n$ we define

$$^*\mathbb{R}_{\Delta x}^n := \{ \Delta x k \,|\, k \in^* \mathbb{Z}^n \}, \tag{241}$$

where $^*\mathbb{Z}$ denotes the set of hyperintegers and $\Delta x k = (\Delta x k_1, \cdots, \Delta x k_n)$. Note that this discretization uses the same discretization size $\Delta x$ in all directions. This is sufficient since we are interested in projections to a classical space, and all classical objects we are interested in can be recovered by this form of

discretization. Note that for a considerable class of data $f :^* \mathbb{R}^n \to^* \mathbb{R}$ (see below) the standard part of the function

$$u(t_M, x) = E\left(f(B^x_{t_M})\right) = \sum_{[\omega]^n_{t_M} \in \Omega^n_M} f\left(B^x_{t_M}([\omega]^n_{t_M})\right) P^n_M([\omega]^n_{t_M}) \qquad (242)$$

satisfies the $n$ dimension heat equation Cauchy problem with initial data $^\circ f$. We prove this using a nonstandard transition density (we could prove this directly, but the following considerations concerning the transition density are useful for our argument later on). Similar as in the case $n = 1$ in general dimension $n \geq 1$ we introduce the density function $p$ defined on $T \times^* \mathbb{R}^n_{\Delta x} \times^* \mathbb{R}^n_{\Delta x}$ by

$$p(t_M, x, y) := \sum_{[\omega]^n_{t_M} \in \Omega^n_M} \delta^n_y\left(B^x_{t_M}([\omega]^n_{t_M})\right) P^n_M([\omega]^n_{t_M}), \qquad (243)$$

where for each $y \in^* \mathbb{R}^n_{\Delta x}$

$$\delta^n_y(z) := \begin{cases} 1 \text{ iff } z = y \\ \\ 0 \text{ iff } z \neq y \end{cases} \qquad (244)$$

denotes the hyperfinite $n$-dimensional Kronecker delta translated by $y$. The standard part of the function $p$ can be computed as a limit similar as in the nonstandard proof of the central limit theorem above. Furthermore, the internal function $u$ has a representation

$$u(t_M, x) = \sum_{y \in^* \mathbb{R}^n} \sum_{[\omega]^n_{t_M} \in \Omega^n_M} f(y) \delta^n_y\left(B^x_{t_M}([\omega]^n_{t_M})\right) P^n_M([\omega]^n_{t_M}), \qquad (245)$$

and this representation may be used to get another proof of the Feynman-Kac formula for $L^1$ data in standard space. Let $\Pi_t$ denote the projection Consider the

$$p(t_M + \Delta t, x, y) := \sum_{[\omega]^n_{t_{M+1}} \in \Omega^n_{M+1}} \delta^n_y\left(B^x_{t_{M+1}}([\omega]^n_{t_{M+1}})\right) P^n_{M+1}([\omega]^n_{t_{M+1}})$$

$$= x + \sum_{\omega(t) \in \{-1,1\}^n} \sum_{j=1}^n \left(\omega_j(t)\sqrt{\Delta t}\right) \mathbf{e}_j P^n_1(\omega(t))$$

$$+ \sum_{\omega(t) \in \{-1,1\}^n} \sum_{[\omega]^n_{t_M} \in \Omega^n_M} \delta^n_y\left(B^{x+\omega(t)\sqrt{\Delta t}}_{t_M}([\omega]^n_{t_M})\right) P^n_M([\omega]^n_{t_M})$$

$$= p(t_M, x, y) + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} p(t_M, x, y) \Delta t + O\left(\Delta t^{3/2}\right),$$

$$(246)$$

and this leads to the conclusion that the standard part of the function $p$ satisfies the heat equation with respect to the variable $x$ (heat equation with $\frac{1}{2}$ times the Laplacian, to be precise). Similarly, the standard part of $p$ satisfies the adjoint heat equation with respect to $y$ (similar proof). Hence the standard part of $p$ is indeed the transition density.

Next let us consider a parabolic system with constant coefficients $a^{ij}_{kl}$, $b^i_{jk}$, and $c^i_j$ on the $n$ torus $\mathbb{T}^n = (\mathbb{R}^n/\mathbb{Z})^n$ of the form

$$\begin{cases} \frac{\partial u_i}{\partial t} = \sum_{j,k,l=1}^n a^{ij}_{kl} \frac{\partial^2 u_j}{\partial x_k \partial x_l} + \sum_{j,k=1}^n b^i_{jk} \frac{\partial u_j}{\partial x_k} + \sum_{j=1}^n c^i_j u_j, \\ \\ u_i(0, x) = f_i\left(0, x + \mathbf{e}_j\right) \end{cases} \qquad (247)$$

for all $\mathbf{e}_j$ and some data $\mathbf{f} = (f_1, \cdots, f_n)^T$. For simplicity of notation we consider the case where the constants of the lower order terms are zero, i.e., $b^i_{jk} = 0$ and $c^i_j = 0$. We try to determine the solution in the form

$$\mathbf{u}(t,x) = \sum_{\alpha \in \mathbb{Z}^n} C_\alpha(t) \exp{(i2\pi\alpha x)}, \tag{248}$$

where

$$C_\alpha(t) = \left(C^1_\alpha(t), \cdots, C^n_\alpha(t)\right)^T \tag{249}$$

are vector-valued functions to be determined. Plugging (248) into (247) we get

$$\frac{\partial C^i_\alpha}{\partial t} = -\sum_{j,k,l=1}^n a^{ij}_{kl} 4\pi^2 \alpha_k \alpha_l C^j, \tag{250}$$

or, in matrix notation (with $A_\alpha = (A^{ij}_\alpha) := \left(\sum_{kl} a^{ij}_{kl} 4\pi^2 \alpha_k \alpha_l\right)$), this is

$$\frac{\partial C_\alpha}{\partial t} = -A_\alpha C_\alpha. \tag{251}$$

Since $A$ is positive definite according to our assumptions we have

$$C_\alpha(t) = \exp{(-A_\alpha t)} \, C_\alpha(0). \tag{252}$$

Moreover $C^i_\alpha$ must be the $\alpha$th Fourier coefficients $f_{i\alpha}$ (with respect to some order of multiindices) of the function $f_i$, i.e.,

$$f_{i\alpha} = \int_{\mathbb{T}^n} f_i(y) \exp{(-i2\pi\alpha y)} \, dy \tag{253}$$

$$f_i(x) = \sum_{\alpha \in \mathbb{Z}^n} f_{i\alpha} \exp{(i2\pi\alpha x)}. \tag{254}$$

Let

$$f_\alpha = (f_{\alpha 1}, \cdots, f_{\alpha n})^T. \tag{255}$$

Hence, we have

$$\mathbf{u}(t,x) = \sum_{\alpha \in \mathbb{Z}^n} \int_{\mathbb{T}^n} \exp{(-A_\alpha t)} \, \mathbf{f}(y) \exp{(i2\pi\alpha(x-y))} \, dy. \tag{256}$$

Next we define

$$\Theta^A(t, x-y) = \sum_{\alpha \in \mathbb{Z}^n} \exp{(-A_\alpha t)} \, \overrightarrow{\exp}{(i2\pi\alpha(x-y))}, \tag{257}$$

where

$$\overrightarrow{\exp}{(i2\pi\alpha(x-y))} = (\exp{(i2\pi\alpha(x-y))}, \cdots, \exp{(i2\pi\alpha(x-y))})^T \tag{258}$$

Next recall that $A_\alpha = Q_\alpha \Lambda_\alpha Q^T_\alpha$ for diagonal $\Lambda_\alpha$ with positive entries $\lambda_i > 0$ and orthogonal $Q$. Hence,

$$A_\alpha = \left(Q_\alpha \Lambda^{1/2}_\alpha Q^T_\alpha Q_\alpha \Lambda^{1/2}_\alpha Q^T_\alpha\right) = \left(Q_\alpha \Lambda^{1/2}_\alpha Q^T_\alpha\right)^2 \tag{259}$$

41

Next define (with $\Delta x = \sqrt{\Delta t}$)

$$B^\Lambda : T \times \Omega_n \to^* \mathbb{R}^n_{\Delta x} \otimes^* \mathbb{R}^n_{\Delta x}$$

$$\left(B^\Lambda(t,\omega)\right)_{ij} = \sum_{i=1}^n \left(\sum_{s<t} \omega_i(s)\sqrt{\Delta t}\right) \lambda_i \delta_{ij}, \qquad (260)$$

where $\delta_{ij}$ denotes the classical Kronecker $\delta$, and for a positive definite matrix $A$ with decomposition $Q\Lambda Q^T$ and $\Lambda = \mathrm{diag}\,(\lambda_i)$ define

$$B^{\sqrt{A}} : T \times \Omega_n \to^* \mathbb{R}^n_{\Delta x} \otimes^* \mathbb{R}^n_{\Delta x}$$

$$B^{\sqrt{A}}(t,\omega) := QB^{\Lambda^{1/2}}(t,\omega)Q^T. \qquad (261)$$

Then we have

$$E\left[\exp\left(iB^{\sqrt{A}}(t,.)\right)\right] \approx \exp\left(-At\right), \qquad (262)$$

i.e.

$$E\left[\exp\left(iB^{\sqrt{A}}(t,.)\right)\right] - \exp\left(-At\right) \qquad (263)$$

equals a matrix with infinitely small entries. Hence, we have the representation

$$\boldsymbol{\Theta}^A(t,x-y) \approx \sum_{\alpha \in \mathbb{Z}^n} E\left[\exp\left(i\alpha B^{\sqrt{A_\alpha}}(t,.)\right) \overrightarrow{\exp}\left(i2\pi\alpha(x-y)\right)\right] \qquad (264)$$

Hence, we have the representation

$$\mathbf{u}(t,x) = \sum_{\alpha \in \mathbb{Z}^n} \int_{\mathbb{T}^n} \exp\left(-A_\alpha t\right) \mathbf{f}(y) \exp\left(i2\pi\alpha(x-y)\right) dy$$

$$\approx \sum_{\alpha \in \mathbb{Z}^n} E\left[\exp\left(i\alpha B^{\sqrt{A_\alpha}}(t,.)\right) \int_{\mathbb{T}^n} \mathbf{f}(y)\exp\left(i2\pi\alpha(x-y)\right) dy\right] \qquad (265)$$

$$= \sum_{\alpha \in \mathbb{Z}^n} E\left[\exp\left(i\alpha B^{\sqrt{A_\alpha}}(t,.)\right) \hat{\mathbf{f}}_\alpha\right].$$

The theorems (2) and (4) then follow by standard arguments. Next we consider systems with variable second order coefficients $x \to a^{ij}_{kl}(x)$. What we need for our construction is global existence and the requirement that for all $1 \le i, j \le n$ the matrices

$$A_\alpha(x) = (A^{ij}_\alpha)(x) := \left(\sum_{kl} a^{ij}_{kl}(x)\alpha_k\alpha_l\right) \qquad (266)$$

are uniformly elliptic in $x \in \mathbb{R}^n$ and for all $\alpha$ where all $\alpha_i \ne 0$. In order to avoid technicalities we assume that for all $\alpha \in \mathbb{Z}^n$ the functions $x \to A_\alpha(x)$ are $C^\infty$, bounded, and with bounded derivatives. We may consider the matrices in (266) on a hyperfinite lattice $^*\mathbb{R}^n_{\Delta x}$ and use the same symbol $A_\alpha = A_\alpha(x)$, where now $x \in^* \mathbb{R}^n_{\Delta x}$. For each $\alpha \in \mathbb{Z}^n$ and each $x \in^* \mathbb{R}^n_{\Delta x}$ consider the positive definite matrix $A_\alpha(x)$ with decomposition $A_\alpha(x) = Q^\alpha_x \Lambda_x Q^{\alpha,T}_x$ and $\Lambda_x = \mathrm{diag}\,(\lambda_{xi})$. Then define the increment

$$\Delta X^{\sqrt{A_\alpha}} : T \times \Omega_n \to^* \mathbb{R}^n_{\Delta x} \otimes^* \mathbb{R}^n_{\Delta x}$$

$$\Delta X^{\sqrt{A_\alpha}}(t,\omega) := Q_x \Delta B^{\Lambda_x^{1/2}}(t,\omega)Q^T_x, \qquad (267)$$

where

$$\Delta B^{\sqrt{\Lambda_x}^{1/2}}(t,\omega) = \sum_{i=1}^n \omega_i(t)\sqrt{\Delta t}\mathbf{e}_i\sqrt{\lambda_{xi}}. \qquad (268)$$

42

Then the $\alpha$ mode $X^{\sqrt{A_\alpha}} : T \times \Omega_n \to^* \mathbb{R}^n_{\Delta x} \otimes^* \mathbb{R}^n_{\Delta x}$ of a generalized process $X^{\otimes,A}(t,.)$ s may be defined by an internal sum of increments $\Delta X^{\sqrt{A_\alpha}}$. A generalized process is then given by

$$X^{\otimes,A}(t,.) := \left( \exp\left( i\alpha X^{\sqrt{A_\alpha}}(t,.) \right) \right)_{\alpha \in \mathbb{Z}^n} \qquad (269)$$

and can be used in order to set up a probabilistic scheme for parabolic systems with variable coefficients.

# References

[1] BACHELIER, L., *Theorie de probabilité*, Gauthier-Villars, Paris (1912).

[2] BRILLOUIN, L., *Science and information theory*, New York (1956).

[3] BECHERER, D., SCHWEIZER, M., *Classical solutions to reactiondiffusion systems for hedging problems with interacting It and point processes*, Ann. Appl. Probab. Volume 15, Number 2 (2005), 1111-1144.

[4] EINSTEIN, A., *Investigations of the theory of Brownian motion*, Dover (1956).

[5] FRIES, CHRISTIAN; KAMPEN, JÖRG: Proxy Simulation Schemes for generic robust Monte Carlo sensitivities, process oriented importance sampling and high accuracy drift approximation (with applications to the LIBOR market model), *Journal of Computational Finance,* Vol. 10, Nr. 2, 97-128, 2007.

[6] Gilkey, P.,*Invariance theory, the Heat equation, and the Atiyah Singer Index Theorem.*, Studies in advanced mathematics, CRC press, (1995).

[7] KAMPEN, J., KOLODKO, A., SCHOENMAKERS, J., *Monte Carlo Greeks for financial products via approximative transition densities*, Siam J. Sc. Comp., vol. 31 , p. 1-22, (2008).

[8] KAMPEN, J., *Global regularity and probabilistic schemes for free boundary surfaces of multivariate American derivatives and their Greeks*, Siam J. Appl. Math. 71, pp. 288-308, (2011).

[9] KRYLOV, N.V., *Lectures on Elliptic and Parabolic Equations in Hölder Spaces,* Graduate Studies in Mathematics, Vol. 12, American Mathematical Society, (1996).

[10] NELSON, E., *Quantum fluctuations*, Princeton University Press, Princeton (1985).

[11] ROUX, S., *Generalized Brownian motion and elasticity*, Journal of Statistical Physics, (1987), Vol. 48, no. 7, p. 201-213.

[12] SABELFELD, K., *Monte Carlo Methods in Boundary Value Problems (Scientific Computation)* , Springer, (1991).

[13] SHANNON, L., *A mathematical theory of communication*, Bell Syst. Tech. Journal, Vol. 27, 379-423, 623-656, (1948).

[14] WIENER, N., *Differential space*, Journal of Mathematical Physics, (1923), p. 131-174.